Instantons in $D = 5$ super-Yang-Mills theory

Master of Science (Dissertation)

Nirina Maurice Hasina Tahiridimbisoa
(nirinamaurice.hasinatahiridimbisoa@students.wits.ac.za)
University of the Witwatersrand

Supervised by: Professor Robert de Mello Koch
(robert@neo.phys.wits.ac.za)
University of the Witwatersrand

February 2014
Submitted to the Faculty of Science Wits University

Declaration

I, the undersigned, hereby declare that the work contained in this M.Sc dissertation thesis is my original work, and that any work done by others or by myself previously has been acknowledged and referenced accordingly. This dissertation is submitted to the University of the Witwatersrand, Johannesburg, in fulfillment of the requirements for the degree of Master of Science. It has not beeen submitted before for any degree or examination in any other university.

Nirina Maurice Hasina Tahiridimbisoa, 18 February 2014

Abstract

One of the key goals of string theory is to provide a unification of general relativity and quantum field theory. In the pursuit of this goal it has become clear that the different string theories that have been discovered so far are all in fact, partial descriptions of a single theory. At strong coupling a new theory, called M-theory, is the correct description. M-theory includes gravitons, M2-branes and M5-branes. Up to now, the correct description of the M5-brane is outstanding. In this project some proposals for this theory are studied.

In particular, there is a proposal that D=5 maximally supersymmetric Yang-Mills theory can be used to provide a description of the world volume physics of the M5-brane. According to this proposal, instantons in D=5 maximally supersymmetric Yang-Mills theory are graviton excitations of the M theory. In this M.Sc dissertation the instanton solutions of D=5 maximally supersymmetric Yang-Mills theory are explored, with the goal of testing the above proposal. The dissertation begins with a review of the uses of instantons in quantum mechanics. In particular, instantons are used to account for tunneling effects within a path integral approach to quantum mechanics. The lifting of ground state degeneracies as well as the estimation of the lifetime of unstable states using instantons is developed. The quantization of gauge theories is reviewed in detail. The relevance of instantons for a semi-classical study of Yang-Mills theory is explained. Finally, the relevance of instantons for $D = 5$ maximally supersymmetric Yang-Mills theory is considered.

Acknowledgements

I would like to express my gratitude to my supervisor Professor Robert de Mello Koch for his funding support and his wonderful supervising. This work is also made possible by the funding support of a half bursary from the African Institute for Mathematical Sciences.

I do not also forget to thank especially my family for being patient during this one year of completing this project.

Furthermore, I would like to thank Stuart Graham, Garreth Kemp and Warren Carlson for having their times for discussions.

Contents

List of Figures

1. Introduction

Unification has been a remarkably fruitful idea. Indeed, ever since Maxwell first synthesized electricity and magnetism into a single theory of electromagnetism and, in the process, explained that light is an electromagnetic wave, unification has been an important guiding principle. Currently, unification is driving the search for a single framework uniting the classical theory of gravity, described in the framework of Einstein's General Relativity, and the remaining forces, described using the framework of quantum field theory. This unified description, dubbed quantum gravity, still appears to be far from completed.

String theory is a promising candidate theory of quantum gravity. Perturbative string theory is obtained through the quantization of a two dimensional superconformal quantum field theory living on the world-sheet of the (one dimensional) fundamental string. This theory has a rich spectrum of spacetime states, rich enough to capture the physics of the all known interactions we see in nature. There are at least five different supersymmetric string theories in ten dimensions. The specific theory one constructs depends on the possible choices we make about the topology and field content of the fundamental string.

Open strings are described by worldsheets that have a boundary. A consequence of this boundary is that the left- and right-moving fields in the worldsheet quantum field theory are identified due to the boundary conditions of the worldsheet. The resulting theory is known as type I string theory. The bosonic spacetime states arising from the open string include a massless spacetime vector field, A_μ, with an $SO(32)$ gauge group.

If the string is closed the worldsheet has no boundary. In this case the left and right moving fields are essentially independent. There are two possible closed string theories known as type IIA and type IIB. These theories differ in the chirality of the left and right moving fields. The spectrum of the type IIA and type IIB string theories both include the massless spacetime states for a scalar field (called the dilaton), an antisymmetric field with two indices (called the Kalb-Ramond 2-form) and a traceless symmetric field with two indices (the graviton). The graviton is the quantum of the gravitational field and its appearance in the spectrum is why string theory is a quantum theory of gravity. There are other massless states that are not shared by type IIA and type IIB string theory. Type IIA contains a 1-form and a 3-form field, and type IIB containing a 0-form, a 2-form and selfdual 4-form field.

The interaction of open strings in type I string theory will create closed strings. These closed strings are in fact those of type IIB theory projected so that they are unoriented. This projected string gives rise to a spacetime dilaton, graviton and 2-form field, so that the type I theory is also a theory of quantum gravity.

The remaining two string theories come from the peculiar choice of taking the left moving fields to be purely bosonic while the right moving fields are supersymmetric. To obtain a consistent theory, we have to set things up so that the left moving and right moving fields live in a different number of spacetime dimensions. To obtain a consistent theory with a sensible interpretation, the bosonic fields are compactified on a 16-dimensional manifold. There are two consistent choices for this compactification which gives rise to the $E_8 \times E_8$ and $SO(32)$ heterotic string theories.

If string theory is to be our theory of everything, then it is a little puzzling that there are five consistent string theories. It would be much more attractive if there was a single unique theory. This difficulty may be resolved by the discovery that the five string theories can be related by a web of dualities. The type IIA and type IIB theories are related under T-duality. What this means is, if the IIA and IIB theories are studied on a spacetime that includes a circle, then the physics of IIA string theory on a small circle is

identical to the physics of IIB string theory on a big circle and vice versa. T-duality is a "big circle-small circle" equivalence. The heterotic string theories are also related under T-duality. Type I string theory at small string coupling is related to the SO(32) heterotic string theory at large coupling. This is known as S-duality. These connections between the different string theories strongly suggests that they share a common origin and hence that there is a more fundamental underlying theory.

The conjectured theory that links these five string theories has been dubbed M-theory. It is an eleven dimensional theory. We do not yet have a fundamental description of M-theory (as we do with string theory in terms of the fundamental string). We can, however, infer a lot about the objects that must arise in M-theory knowing that it must contain the five string theories and eleven dimensional supergravity. Indeed, the field content of eleven dimensional supergravity, implies that the massless bosonic fields in M-theory must be a graviton and a 3-form. Eleven dimensional supergravity also contains stable 2-brane and 5-brane solutions, suggesting that M-theory contains appropriate M2-branes and M5-branes.

The M5-brane has defied description. Recent speculative conjectures suggest that the M5-brane theory on a circle is dual to the five dimensional super-Yang-Mills theory that arises as the low energy action of multiple D4-branes. This is a $4+1$-dimensional theory. The M5-brane theory should be a $5+1$-dimensional theory. How does the missing dimension emerge from the five dimensional super-Yang-Mills theory? Instantons play an important role in this conjecture [1]. Indeed, they are supposed to encode momentum modes of the emergent dimension. Instantons are solutions in four dimensional Euclidean Yang-Mills theory which are characterized by a selfdual field strength. They were discovered 35 years ago and have played an important role in understanding quantum tunnelling in four dimensional Yang-Mills, and in understanding supersymmetric gauge theories. For our present application to five dimensional Yang-Mills theory, instantons are solutions in the four spatial dimensions and can evolve in the time direction. Very little is known about the dynamics of instantons, in this setting.

Notice that although string theory has played a prominent role is the specific motivation for our problem, we could easily motivate our study as an exercise in quantum field theory, or in dualities. We do not know whether string theory will ultimately give a useful or sensible description of anything in physics, so it is important to bear this comment in mind.

The instantons we study have a slow roll instability in the soliton size. If the static soliton is given a small perturbation, it will either shrink or expand at a constant rate until it hits zero size or expands indefinitely. This pathological behaviour seems to conflict with the purported role of instantons as momentum modes for an emergent dimension. However, instantons can be stabilized with the addition of a potential. In this case, the solutions must also be given an appropriate rotation to stabilize against collapse from the potential, and in doing so they gain an electric charge. These charged solitons, known as dyonic instantons in five dimensional Yang-Mills, are stable to perturbations.

The space of all physically different instanton solutions is known as the moduli space. The low energy dynamics of topological solitons can generally be approximated by geodesic motion through the moduli space. Intuitively, the static solitons are the lowest energy configurations in each topological sector, and giving them a small velocity will only increase their energy slightly. These configurations will then always stay close to the minimum energy configurations by energy conservation. The moduli space can be thought of as the floor of a potential valley in the space of all solutions to the Yang-Mills equations of motion, and the evolution of slow moving solitons must stay near the bottom of this potential valley. So long as the evolution does not ascend too far up the valley, the motion can be approximated by motion exactly along the valley floor, or as the evolution through snapshots of the static solitons in the moduli space. For dyonic instantons, each solution has a different electric charge, which induces a potential on the moduli space. So long as this potential is shallow compared to the surrounding valley,

the moduli space approximation is still valid [2]. Instanton solutions can be found using a technique known as the ADHM construction which recasts the problem of solving the selfdual field equation into a set of non-linear algebraic equations [3]. This technique is used repeatedly in calculations involving instantons, as it generally allows us to turn all expressions involving derivatives and integrals into purely algebraic expressions. The ADHM construction also provides a natural coordinate system on the moduli space of instantons, at least when an explicit parameterization of the ADHM data is known.

In chapter 2 we introduce instantons in the simplest possible setting: quantum mechanics. We explain how instantons dominate the semiclassical limit after the path integral has been continued to imaginary time. We also explain how the analytically continued path integral is in fact computing properties of the ground state. In chapter 3 we review the quantization of Yang-Mills theories. Although this chapter is not strictly needed to understand the instanton physics we will explore, it is included for completeness. Chapter 4 then considers Euclidean Yang-Mills theory. This chapter develops instantons in a quantum field theory setting. Chapter 5 is concerned with reviewing material relevant to the conjectured link to the physics of the M5-brane. Conclusions are collected in chapter 6.

This dissertation reviews a large amount of material. For that reason we recommend some further sources in the vast available literature. The mathematical techniques used to study instantons as well as further applications can be found in [4]. For further background on quantum field theory the textbook [5] is a modern and detailed introduction. Finally, worldsheet string theory as well as a few modern topics are treated in detail in [6, 7].

2. Instantons in Quantum Mechanics

2.1 Transition amplitudes and path integrals

We will develop the basic approach in the simple setting of quantum mechanics. We focus on the ground state wavefunction and ground state energy of the system. A key ingredient is *quantum tunnelling* [8]. Quantum tunnelling is a non classical physics phenomenon. It only occurs in quantum mechanics and allows a particle to tunnel through a potential barrier. To illustrate this idea, consider the following figure.

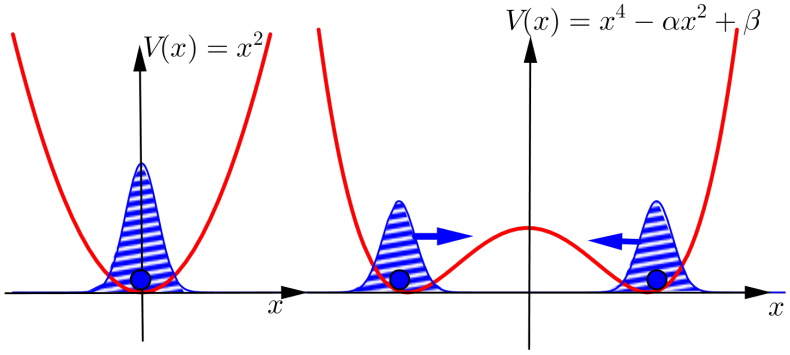


Figure 2.1: Particles located at the minima of the potential. The wavefunctions of the particles are indicated by striped functions.

In the above figure, the particles are represented by the small blue balls. The particles were placed at their most probable positions in the ground state. The arrows represent the possible quantum tunnelling events.

To study properties of the ground states we will see that it is useful to start by considering the transition amplitude for a particle to move from x_i at time t_i to x_f at time t_f ($t_f > t_i$), [9, 5]. The path integral evaluates this transition amplitude by summing over all possible paths the particle could have used to make this transition.

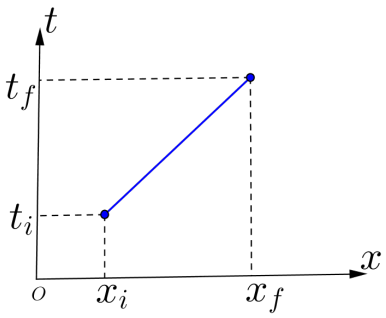


Figure 2.2: A spacetime diagram showing one possible path that could be followed by a particle. The path integral performs a sum over all possible paths.

The transition amplitude is

$$\langle x_f, t_f | x_i, t_i \rangle, \tag{2.1.1}$$

which can be written as

$$\langle x_f | e^{-iHT/\hbar} | x_i \rangle, \text{ where } T = t_f - t_i. \tag{2.1.2}$$

In terms of a path integral

$$\langle x_f | e^{-iHT/\hbar} | x_i \rangle = N \int [dx] e^{iS/\hbar} \tag{2.1.3}$$

where H is the Hamiltonian of the system, N is a normalisation constant, S is the classical action of our system and the integral runs over all possible paths. To be concrete we consider a particle moving under the influence of an external potential $V(x)$. The classical action takes the form

$$S = \int_{t_i}^{t_f} dt \left[\frac{1}{2} \left(\frac{dx}{dt} \right)^2 - V(x) \right] = \int_{-\frac{T}{2}}^{\frac{T}{2}} dt \left[\frac{1}{2} \left(\frac{dx}{dt} \right)^2 - V(x) \right]. \tag{2.1.4}$$

To study the ground state we will that that it is useful to consider the change of variable

$$t = -it_E \;\Rightarrow\; T = -iT_E; \tag{2.1.5}$$

under which

$$S = i \int_{-\frac{T_E}{2}}^{\frac{T_E}{2}} dt_E \left[\frac{1}{2} \left(\frac{dx}{dt_E} \right)^2 + V(x) \right], \tag{2.1.6}$$

$$= iS_E. \tag{2.1.7}$$

S_E is called the *Euclidean action*. The Euclidean action is the analytic continuation of the original real time action. Equation (2.1.3) can be rewritten in terms of the Euclidean action and the Euclidean time t_E as

$$\langle x_f | e^{-HT_E/\hbar} | x_i \rangle = N \int [dx] e^{-S_E/\hbar}. \tag{2.1.8}$$

Consider the left hand side of equation (2.1.8). Inserting a complete set of energy eigenstates, we have

$$\langle x_f | e^{-HT/\hbar} | x_i \rangle = e^{-E_o T/\hbar} \left[\langle x_f | 0 \rangle \langle 0 | x_i \rangle + \langle x_f | 1 \rangle \langle 1 | x_i \rangle e^{-(E_1 - E_o)T/\hbar} + \ldots \right]. \tag{2.1.9}$$

In the limit $T \to \infty$, this last expression for the transition amplitude can, with exponential accuracy, be approximated as

$$\langle x_f | e^{-HT/\hbar} | x_i \rangle = e^{-E_o T/\hbar} \langle x_f | 0 \rangle \langle 0 | x_i \rangle. \tag{2.1.10}$$

Combining this last equation with the equation (2.1.8), we have

$$\langle x_f | e^{-HT/\hbar} | x_i \rangle = N \int [dx] e^{-S_E/\hbar}, \tag{2.1.11}$$

$$\underset{T \to \infty}{=} e^{-E_o T/\hbar} \langle x_f | 0 \rangle \langle 0 | x_i \rangle, \tag{2.1.12}$$

where we integrate over paths that obey the boundary conditions

$$x[-T/2] = x_i \text{ and } x[T/2] = x_f.$$

This is a key result since it shows very directly how properties of the ground state can be computed using a path integral.

2.2 Steepest descent saddle point approximation

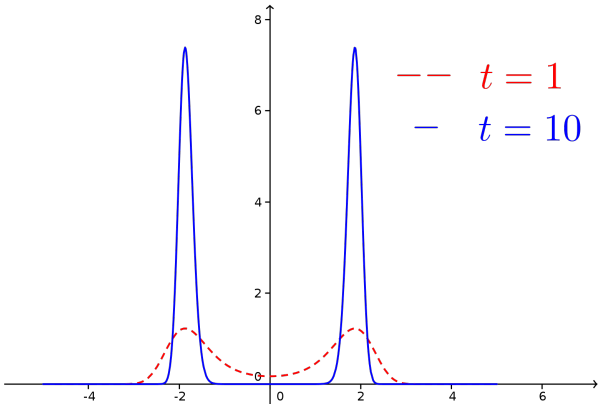


Figure 2.3: Profile of functions $e^{-t[\frac{1}{2.5^2}(x^2-3.5)^2-0.2]}$

Steepest descent is a method used to approximate the value of an integral. The fundamental idea is that the value of the integral should be dominated by contributions from the neighborhoods of points where the integrand peaks, [10, 11]. The integral we consider is

$$I = \int_\zeta dz e^{th(z)}, \tag{2.2.1}$$

where $z = x + iy$, $t \gg 0$ and $h(z)$ is a complex valued function

$$h(z) = \phi(x, y) + i\psi(x, y).$$

Note that we require t is large which ensures that the integral is dominated by the peaks of the integrand. See Figure 2.3 for an illustration of the impact of small and large values of t. The function $h(z)$ is assumed to be *analytic* (holomorphic), which implies the Cauchy-Riemann equations are satisfied

$$\begin{cases} \frac{\partial \phi}{\partial x} &= \frac{\partial \psi}{\partial y}, \\ \frac{\partial \phi}{\partial y} &= -\frac{\partial \psi}{\partial x} \end{cases}.$$

Start by choosing a path ζ_o such that $\psi(x, y)$ is a constant along ζ_o. Since the integrand is not wiggling along this path, the dominant contribution is from the peak of $e^{t\phi}$ which is given by the peak of $\phi(x, y)$. If ϕ changes rapidly along this path, steepest descent will give an accurate answer. This is indeed the case. Noting that ϕ changes most rapidly along the path with tangent given by $\vec{\nabla}\phi$, and that if

$$\begin{aligned} \vec{X}_2 = \vec{X}_1 + \delta\vec{X} &= \vec{X}_1 + \epsilon\vec{\nabla}\phi, \\ \Rightarrow \psi(\vec{X}_2) &= \psi(\vec{X}_1 + \epsilon\vec{\nabla}\phi), \\ &= \psi(\vec{X}_1) + \epsilon\vec{\nabla}\phi \cdot \vec{\nabla}\psi, \\ &= \psi(\vec{X}_1). \end{aligned}$$

Thus ψ is constant along the path for which ϕ varies most rapidly. Note that the passage to the last line above used the Cauchy-Riemann equations. Finally, we can summarize our discussion by saying

that the tangent of the path ζ_o at a certain point, is parallel to $\vec{\nabla}\phi$ at that point and the approximated value of the integral (2.2.1) along the path ζ_o for large t is

$$I = \int_{\zeta_o} dX e^{t\phi(X)}, \tag{2.2.2}$$

$$\approx \int_{\zeta_o} dX \exp\left(t\phi_{\max} + t\phi''(X_{\max})\frac{X^2}{2}\right), \tag{2.2.3}$$

$$= \int dX \exp\left(t\phi_{\max} - t|\phi''|\frac{X^2}{2}\right), \tag{2.2.4}$$

$$= e^{t\phi_{\max}}\sqrt{\frac{2\pi}{t\phi''(X_{\max})}}. \tag{2.2.5}$$

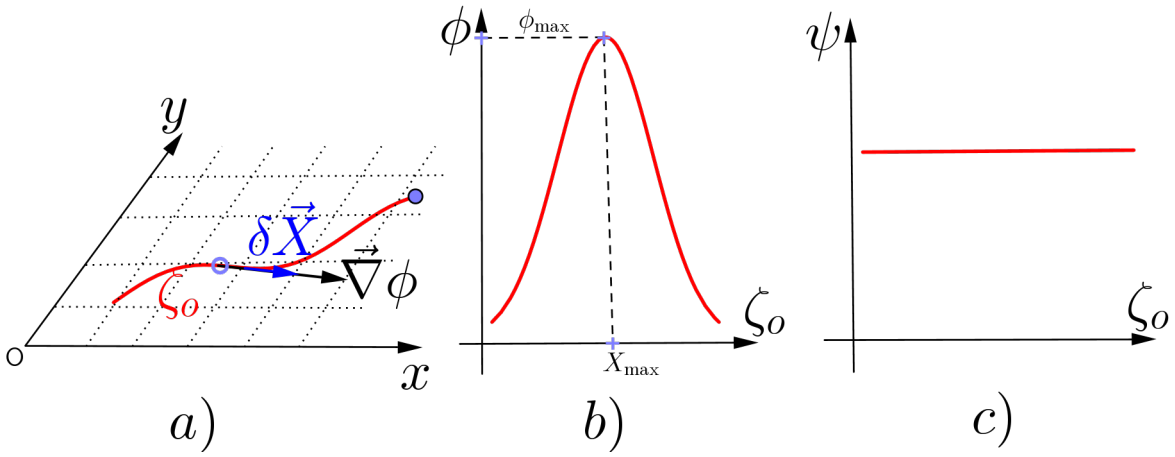


Figure 2.4: a) An example of a path ζ_o in the complex plane, b) the profile of ϕ along the path ζ_o and c) the profile of ψ along ζ_o.

Now use steepest descent to evaluate the path integral in (2.1.11). Steepest descent will be valid in the $\hbar \to 0$ limit. The first step is to find an extremum of the action S_E. Variations around this path vanish to first order. Thus

$$S_E = \int_{\frac{T}{2}}^{\frac{T}{2}} dt \left[\frac{1}{2}\left(\frac{dx}{dt}\right)^2 + V(x)\right],$$

$$\Rightarrow \ \delta S_E = 0 \ \Leftrightarrow \ 0 = \int_{\frac{T}{2}}^{\frac{T}{2}} dt \left[\left(\frac{dx}{dt}\right)\delta\left(\frac{dx}{dt}\right) + \frac{dV(x)}{dx}\delta x\right],$$

$$0 = \cancel{\left[\left(\frac{dx}{dt}\right)\delta x\right]_{x_i}^{x_f}} + \int_{\frac{T}{2}}^{\frac{T}{2}} dt \left[-\left(\frac{d^2x}{dt^2}\right) + \frac{dV(x)}{dx}\right]\delta x.$$

The first term in the last equation vanishes because the end points (x_i and x_f) are fixed for all possible paths $x + \delta x$ which implies that δx vanishes at the endpoints. Finally, we have

$$\frac{d^2x}{dt^2} = \frac{dV(x)}{dx}. \tag{2.2.6}$$

The equation (2.2.6) is called the Euler-Lagrange equation. This equation is the equation of motion for a particle moving in the *inverted potential* $-V(x)$.

For simplicity, assume that S_E has only one extremum, given by the path $\tilde{x}(t)$. For a general path $x(t)$, we write

$$x(t) = \tilde{x}(t) + \sum_n C_n x_n(t), \tag{2.2.7}$$

where $\{x_n\}_{n\in\mathbb{N}}$ is a complete set of orthonormal functions,

$$\int_{-T/2}^{T/2} dt x_n(t) x_m(t) = \delta_{nm}. \tag{2.2.8}$$

To ensure that $x(t)$ satisfies the correct boundary conditions, we require

$$x_n(-T/2) = x_n(T/2) = 0, \quad \text{so that} \tag{2.2.9}$$

$$x(-T/2) = \tilde{x}(-T/2) = x_i, \text{ and } x(T/2) = \tilde{x}(T/2) = x_f. \tag{2.2.10}$$

Thus the integral over all possible paths becomes an integral over C_n. The integral measure is now defined as

$$[dx] = \prod_n \frac{dC_n}{\sqrt{2\pi\hbar}}. \tag{2.2.11}$$

The factor $\frac{1}{\sqrt{2\pi\hbar}}$ is chosen to simplify our calculation. Note that we still have not specified the normalization N in (2.1.11).

In the following we will expand S_E up to the second order in C_n, which is good enough for us to evaluate the path integral using the method of steepest descent. Indeed the $\hbar \to 0$ limit will suppress the contribution of higher orders in C_n in the path integral and enhance the peak at $\tilde{x}(t)$. Expanding S_E to second order in C_n yields

$$\begin{aligned}
S_E[x] &= S_E[\tilde{x} + \textstyle\sum_n C_n x_n], \\
&= \int_{-T/2}^{T/2} dt \left\{ \frac{1}{2}\left[\frac{d\tilde{x}}{dt} + \sum_n C_n \frac{dx_n}{dt}\right]^2 + V[\tilde{x} + \textstyle\sum_n C_n x_n] \right\}, \\
&= S_E[\tilde{x}] + \int_{-T/2}^{T/2} dt \left\{ \frac{d\tilde{x}}{dt} \sum_n C_n \frac{dx_n}{dt} + \frac{1}{2}\sum_{n,m} C_n C_m \frac{dx_m}{dt}\frac{dx_n}{dt} + \left.\frac{dV}{dx}\right|_{\tilde{x}} \sum_n C_n x_n \right. \\
&\quad \left. + \left.\frac{d^2V}{dx^2}\right|_{\tilde{x}} \sum_{n,m} \frac{C_n C_m}{2} x_n x_m + o(C_n^3) \right\}, \\
&= S_E[\tilde{x}] + \int_{-T/2}^{T/2} dt \left\{ \sum_n C_n \left[\frac{d\tilde{x}}{dt}\frac{dx_n}{dt} + V'(\tilde{x}) x_n\right] + \sum_{n,m} \frac{C_n C_m}{2}\left[\frac{dx_n}{dt}\frac{dx_m}{dt} + V''(\tilde{x}) x_n x_m\right] \right\}.
\end{aligned}$$

Note that

$$\begin{aligned}
\int_{-T/2}^{T/2} dt \sum_n C_n \left[\frac{d\tilde{x}}{dt}\frac{dx_n}{dt} + V'(\tilde{x}) x_n\right] &= \left[\sum_n C_n \frac{d\tilde{x}}{dt} x_n\right]_{-T/2}^{T/2} + \int_{-T/2}^{T/2} dt \sum_n C_n \left[-\frac{d^2\tilde{x}}{dt^2} + V'(\tilde{x})\right] x_n, \\
&= 0.
\end{aligned}$$

This integral vanishes after using the Euler-Lagrange equation in (2.2.6) and the boundary conditions in (2.2.9).

In what follows we will use the boundary conditions in (2.2.9), the orthogonality condition in (2.2.8) and have chosen the x_n to obey

$$\left[-\frac{d^2}{dt^2}+V''(\tilde{x})\right]x_m=\lambda_m x_m.$$

λ_n are the eigenvalues of the operator $\left[-\frac{d^2}{dt^2}+V''(\tilde{x})\right]$. First consider

$$\begin{aligned}\int_{-T/2}^{T/2}dt\sum_{n,m}\frac{C_nC_m}{2}\left[\frac{dx_n}{dt}\frac{dx_m}{dt}+V''(\tilde{x})x_nx_m\right]&=\left[\sum_{n,m}\frac{C_nC_m}{2}\frac{dx_n}{dt}x_m\right]_{-T/2}^{T/2}\\&\quad+\int_{-T/2}^{T/2}dt\sum_{n,m}\frac{C_nC_m}{2}x_n\left[-\frac{d^2}{dt^2}+V''(\tilde{x})\right]x_m,\\&=\int_{-T/2}^{T/2}dt\sum_{n,m}\frac{C_nC_m}{2}x_n\lambda_nx_m.\end{aligned}$$

Thus, using orthogonality of the $x_n(t)$, we have

$$\int_{-T/2}^{T/2}dt\sum_{n,m}\frac{C_nC_m}{2}\left[\frac{dx_n}{dt}\frac{dx_m}{dt}+V''(\tilde{x})x_nx_m\right]=\sum_n\frac{C_n^2}{2}\lambda_n.$$

We can now write, correct up to second order in C_n

$$S_E[x(t)]=S_E[\tilde{x}(t)]+\sum_n\frac{C_n^2}{2}\lambda_n+O(C_n^3). \tag{2.2.12}$$

The path integral we are computing reduces to a Gaussian integral. This is easily performed to obtain

$$N\int[dx]e^{-S_E/\hbar}=Ne^{-S_E[\tilde{x(t)}]/\hbar}\int\prod_n\frac{dC_n}{\sqrt{2\pi\hbar}}e^{-\frac{C_n^2}{2\hbar}\lambda_n}, \tag{2.2.13}$$

$$=\frac{Ne^{-S_E[\tilde{x}]/\hbar}}{\sqrt{\prod_n\lambda_n}}, \tag{2.2.14}$$

$$=\frac{Ne^{-S_E[\tilde{x}]/\hbar}}{\sqrt{\det\left(-\frac{d^2}{dt^2}+V''(\tilde{x})\right)}}. \tag{2.2.15}$$

This final result is a non-perturbative answer because $e^{-S_E[\tilde{x}(t)]/\hbar}$ can't be expanded in a series using positive powers of $\hbar$. Working to higher order in C_n will generate corrections that come with positive powers of $\hbar$.

2.2.1 Remarks.

- Our result is obtained assuming there is only one peak of the Euclidean action. However we can generalize our result when there is more than one peak. We simply sum the contribution of each of the additional peaks.
- The second remark is that our result in (2.2.15) holds for any $V(x)$. Thus it is natural to check the validity of our calculations by testing (2.2.15) for the simple harmonic oscillator. Of course, we perform this check because we already know the energy spectrum for a simple harmonic oscillator.

2.3 Example 1: Harmonic oscillator

The aim of this section is to recover the ground state energy of the harmonic oscillator, by using path integral techniques. The potential for the harmonic oscillator is

$$V = \frac{1}{2}\omega^2 x^2. \tag{2.3.1}$$

Thus, the corresponding Euclidean action is

$$S_E = \int_{-T/2}^{T/2} dt L_E = \int_{-T/2}^{T/2} dt \left[\frac{1}{2}\left(\frac{dx}{dt}\right)^2 + \frac{1}{2}\omega^2 x^2\right]. \tag{2.3.2}$$

Our first step is to find the path $\tilde{x}(t)$ that extremizes S_E, by solving the equation of motion. We need to solve

$$\frac{d^2x}{dt^2} = -\left[\frac{-dV}{dx}\right] = \omega^2 x. \tag{2.3.3}$$

The general solution is

$$\tilde{x}(t) = Ae^{\omega t} + Be^{-\omega t}, \tag{2.3.4}$$

where A, B are constants of integration. To find these constants, use the boundary conditions

$$\tilde{x}(-T/2) = 0 \;\Rightarrow\; Ae^{-\omega T/2} + Be^{\omega T/2} = 0, \tag{2.3.5}$$

$$\tilde{x}(T/2) = 0 \;\Rightarrow\; Ae^{\omega T/2} + Be^{-\omega T/2} = 0. \tag{2.3.6}$$

Hence, equations (2.3.5) and (2.3.6) yield

$$A = B = 0 \;\Rightarrow \tilde{x}(t) = 0. \tag{2.3.7}$$

For this solution

$$S_E[\tilde{x}(t)] = 0, \tag{2.3.8}$$

and (2.1.11) and (2.1.12) become

$$\langle x_f|0\rangle\langle 0|x_i\rangle e^{-E_o T/\hbar} = \frac{N}{\sqrt{\det(-\partial_t^2 + \omega^2)}}. \tag{2.3.9}$$

Now, we want to compute the determinant under the square root of (2.3.9). Towards this end, consider

$$(-\partial_t^2 + \omega^2)x_n = \lambda_n x_n, \tag{2.3.10}$$

with the boundary conditions $x_n(\pm T/2) = 0$. We will see that it is useful to study

$$D = \frac{\det[-\partial_t^2 + V^{(1)} - \lambda]}{\det[-\partial_t^2 + V^{(2)} - \lambda]} = \frac{\prod_i(\lambda - \lambda_i^{(1)})}{\prod_j(\lambda - \lambda_j^{(2)})}, \tag{2.3.11}$$

$$= \frac{\psi_\lambda^{(1)}(T/2)}{\psi_\lambda^{(2)}(T/2)}. \tag{2.3.12}$$

The equality shown in (2.3.12) is highly nontrivial and needs to be proved. First, notice that whenever λ is an eigenvalue, so that $\psi_\lambda^{(1)}(t)$ and $\psi_\lambda^{(2)}(t)$ solve (2.3.10), it follows that $\psi_\lambda^{(1)}(T/2)$ and $\psi_\lambda^{(2)}(T/2)$

have a zero. Thus, (2.3.11) and (2.3.12) have the same poles and zeros in λ, implying that the two can differ at most by a constant factor. This is a consequence of a well known theorem in complex analysis that says that a meromorphic is determined, up to an overall constant factor, by the positions of its zeros and poles. We now argue that this constant factor is 1. We can rewrite (2.3.11) and (2.3.12) as

$$\frac{\det[-\partial_t^2 + V^{(1)} - \lambda]}{\psi_\lambda^{(1)}(T/2)} = \frac{\det[-\partial_t^2 + V^{(2)} - \lambda]}{\psi_\lambda^{(2)}(T/2)}, \tag{2.3.13}$$

$$= g(\lambda). \tag{2.3.14}$$

The left hand side of (2.3.13) depends only on $V^{(1)}$ and not on $V^{(2)}$ while the right hand side depends on $V^{(2)}$ and not on $V^{(1)}$. Obviously, both sides depend on λ but are independent of $V^{(1)}$ and $V^{(2)}$. The above statements can be summarized by writing

$$\det[-\partial_t^2 + V^{(1)} - \lambda] = g(\lambda)\psi_\lambda^{(1)}(T/2). \tag{2.3.15}$$

To determine $\psi_\lambda^{(i)}(t)\Big|_{i=1,2}$, we need to solve

$$(-\partial_t^2 + V^{(i)} - \lambda)\psi_\lambda^{(i)}(t) = 0, \tag{2.3.16}$$

$$\text{such that } \begin{cases} \psi_\lambda^{(i)}(-T/2) & = 0, \\ \frac{\partial \psi_\lambda^{(i)}}{\partial t}\Big|_{-T/2} & = 1 \end{cases}. \tag{2.3.17}$$

The last condition above sets an arbitrary normalization. For the simple harmonic oscillator we have

$$V^{(1)} = \omega^2. \tag{2.3.18}$$

Hence, we find from equation (2.3.15) that

$$\frac{N}{\sqrt{\det(-\partial_t^2 + \omega^2)}} = \frac{N}{\sqrt{g(0)\psi_0^{(1)}(T/2)}}. \tag{2.3.19}$$

We will now choose the normalization constant N to get a simple final expression. Choose the normalization constant N such that

$$\pi\hbar N^2 = g(0) = \frac{\det(-\partial_t^2 + \omega^2)}{\psi_0^{(2)}(T/2)}. \tag{2.3.20}$$

It then follows that equation (2.3.9) can be written as

$$\langle x_f|0\rangle\langle 0|x_i\rangle e^{-E_o T/\hbar} = \frac{1}{\sqrt{\pi\hbar\psi_0^{(1)}(T/2)}}. \tag{2.3.21}$$

(2.3.21) tells us that our problem of finding $\det(-\partial_t^2 + \omega^2)$ is reduced to finding the analytic form of $\psi_0^{(1)}(t)$ and then the value $\psi_0^{(1)}(T/2)$. $\psi_0^{(1)}(t)$ satisfies the equation

$$(-\partial_t^2 + \omega^2)\psi_0^{(1)}(t) = 0, \tag{2.3.22}$$

(2.3.22) is a second order linear differential equation which has general solution

$$\psi_0^{(1)}(t) = \alpha e^{\omega t} + \beta e^{-\omega t}, \tag{2.3.23}$$

where α and β are arbitrary constants. Using the boundary conditions in (2.3.17) we find the unique solution

$$\psi_0^{(1)}(t) = \frac{1}{2\omega}\left(e^{\omega(t+T/2)} - e^{-\omega(t+T/2)}\right). \tag{2.3.24}$$

Thus

$$\psi_0^{(1)}(T/2) = \frac{\sinh(\omega T)}{\omega}, \tag{2.3.25}$$

$$\psi_0^{(1)}(T/2) \underset{T\to\infty}{=} \frac{e^{\omega T}}{2\omega}. \tag{2.3.26}$$

This last equation implies

$$\frac{1}{\sqrt{\pi\hbar\psi_0^{(1)}(T/2)}} = \sqrt{\frac{2\omega}{\pi\hbar}}e^{-\omega T/2}. \tag{2.3.27}$$

Equations (2.3.21) and (2.3.27) plus the identification of the argument of the exponential factors yield

$$\frac{\omega T}{2} = \frac{E_o T}{\hbar}; \;\Rightarrow E_o = \frac{\hbar\omega}{2}, \tag{2.3.28}$$

which is the correct ground state energy for a simple harmonic oscillator. We can recover further energy eigenvalues E_n of the harmonic oscillator by noting

$$\frac{1}{\sqrt{\pi\hbar\psi_0^{(1)}(T/2)}} = \frac{1}{\sqrt{\frac{\pi\hbar}{2\omega}e^{\omega T}(1-e^{-2\omega T})}}, \tag{2.3.29}$$

$$= \sqrt{\frac{2\omega}{\pi\hbar}}e^{-\frac{\omega T}{2}}[1-e^{-2\omega T}]^{-\frac{1}{2}}, \tag{2.3.30}$$

$$= \sqrt{\frac{2\omega}{\pi\hbar}}e^{-\frac{\omega T}{2}}\left[1+\sum_{n\geq 1}a_n e^{-2n\omega T}\right]. \tag{2.3.31}$$

Above we did not specify the explicit expression of the factors a_n. In fact they only need to be non-zero since we read energies from the argument of the exponential. The coefficient of the exponential, a_n, is proportional to the value of the wavefunction at $x = 0$. Consequently, the formula (2.3.31) will receive contributions from all energy eigenstates whose position space wave functions are non-zero at $x = 0$. Hence, after identifying the arguments of the exponential factors as we did above, the relevant energy eigenvalues which belong to wave functions that are non-zero at $x = 0$, are

$$E_{2n} = (2n + \frac{1}{2})\hbar\omega. \tag{2.3.32}$$

The first eight wavefunctions of the simple harmonic oscillator have the form

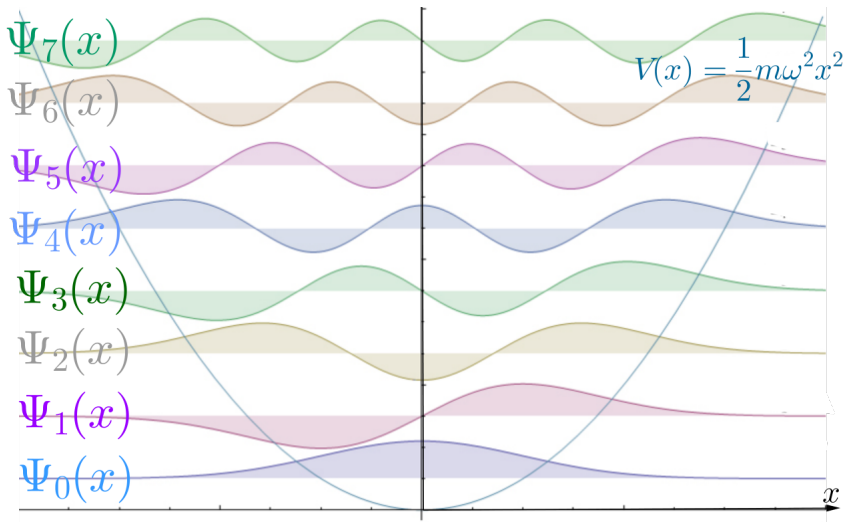


Figure 2.5: Wavefunctions of the harmonic oscillator.

Figure 2.5 tells us that the wavefunctions $\Psi_1, \Psi_3, \Psi_5, \Psi_7$ are odd and they vanish at $x = 0$. Furthermore the remaining wavefunctions are even and do not vanish at $x = 0$. The answer (2.3.32) beautifully reproduce the energy eigenvalues of the contributing energy eigenstates.

2.4 Example 2: Double well potential

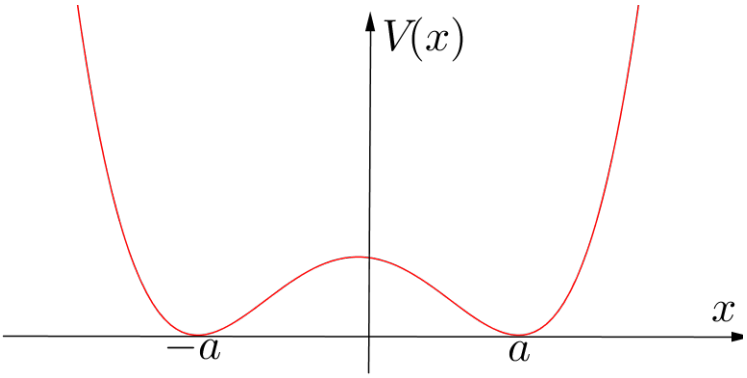


Figure 2.6: $V(x) = x^4 - \alpha x^2 + \beta$ where $\alpha, \beta > 0$.

In this example we are going to consider a potential which has the form given in Figure 2.6 and satisfies $V''(\pm a) = \omega^2$. Since this potential has two extrema at $x_1 = -a$ and $x_2 = a$, we expect to have many extrema of the Euclidean action. Consider $\tilde{x_j}(t)$, the paths which extremize the action, where j labels distinct extrema. In this case we have

$$\langle x_f|0\rangle\langle 0|x_i\rangle e^{-E_o T/\hbar} = \sum_j \frac{N e^{-S_E[\tilde{x}_j]}}{\sqrt{\det(-\partial_t^2 + V''[\tilde{x}_j])}}. \tag{2.4.1}$$

The four possible boundary conditions for x_i and x_f are

1) $x_i = a$ and $x_f = a$,

2) $x_i = -a$ and $x_f = -a$,

3) $x_i = a$ and $x_f = -a$ and

4) $x_i = -a$ and $x_f = a$.

In order to find the paths $\tilde{x_j}(t)$ that extremize the Euclidean action, we need to solve the equation of motion

$$\frac{d^2x}{dt^2} = V'(x). \tag{2.4.2}$$

This differential equation is tough to solve analytically for general potentials $V(x)$. Thus, we will develop another way of finding the paths which yield the extrema of S_E. We can solve the problem if we focus on the "conserved energy" of the particle. The conserved energy is given by

$$E = \frac{1}{2}\left(\frac{dx}{dt}\right)^2 - V(x). \tag{2.4.3}$$

Indeed

$$\frac{dE}{dt} = \frac{dx}{dt}\frac{d^2x}{dt^2} - \frac{dV}{dt} = \frac{dx}{dt}\left(\frac{d^2x}{dt^2} - \frac{dV}{dx}\right),$$

which vanishes by the Euclidean equation of motion. The solutions we study have

$$E = 0. \tag{2.4.4}$$

To see this, note that we start from a configuration with

$$\left.\frac{dx(t)}{dt}\right|_{t=t_i} = 0,$$

and the position of the particle is at $x(t_i) = a$ (or $-a$) for which $V(\pm a) = 0$. Hence the total energy vanishes at this initial time, and consequently the energy of the particle vanishes at any time.

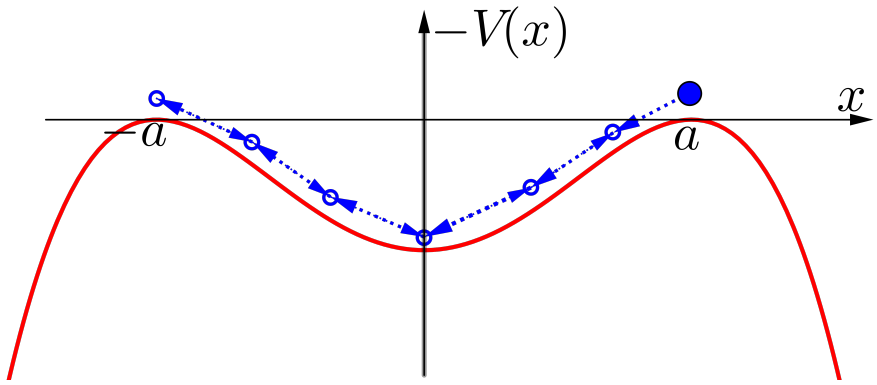


Figure 2.7: Particle located at the position $x = a$. The blue empty circles represent the "motions" of the particles which reproduces tunnelling.

Consider first the case of boundary conditions where $x_i = a$ and $x_f = a$. Most of the time the particle is at rest and at the lowest potential $V(a) = 0$ at $x = a$. Hence the total energy vanishes. Consider equation (2.4.4)

$$E = \frac{1}{2}\left(\frac{dx}{dt}\right)^2 - V(x) = 0 \Leftrightarrow \frac{dx}{dt} = \pm\sqrt{2V(x)}. \tag{2.4.5}$$

Thus, we can write

$$\frac{dx}{dt} = -\sqrt{2V(x)}, \tag{2.4.6}$$

$$\Rightarrow dt = -\frac{dx}{\sqrt{2V(x)}}, \tag{2.4.7}$$

$$\Rightarrow t = t_1 - \int_0^x \frac{dy}{\sqrt{2V(y)}} \tag{2.4.8}$$

where we only focus on the negative root. In Figure 2.8 we sketch the solution given by (2.4.8). Figure 2.8 is most easily understood from equation (2.4.6). We start at $t = -\infty$ from $x = a + \epsilon$. At $x = a$ $V(a) = 0$ so that, according to (2.4.6) our function has an infinitesimally small and negative derivative with respect to t. After an infinite time $x(t)$ decreases and passes through the origin $x = 0$ at $t = t_1$. At this point $\frac{dx}{dt}$ is minimum. It then start to increase and approaches 0 at $t \to \infty$ and $x \to -a$. By convention, this solution is called an anti-instanton. We describe the figure as an anti-instanton of size d at the time t_1. The name instanton is reserved for the opposite solution which interpolates from $x_i = -a$ to $x_f = a$.

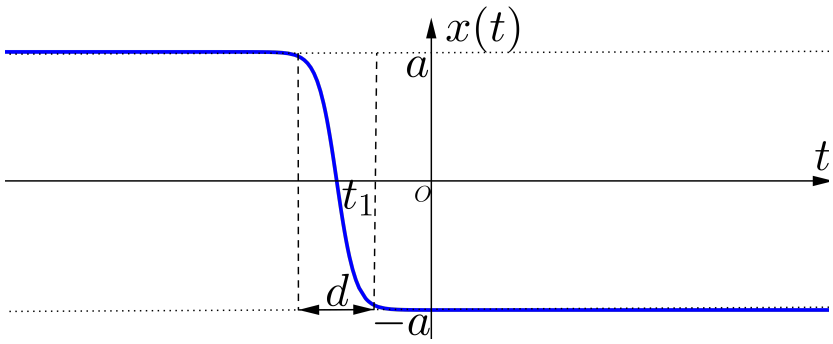


Figure 2.8: Diagram of $x(t)$ showing one anti-instanton for the case $x_i = a$ and $x_f = -a$.

The figure which corresponds to an instanton of size d at a time t_2 is shown in Figure 2.9. The equation describing the instanton is

$$\frac{dx}{dt} = \sqrt{2V(x)}. \tag{2.4.9}$$

Note that we can rederive the equation of motion (2.4.2) by differentiating the equation of the instanton, (2.4.9). Indeed, we have

$$\begin{aligned}\frac{dx}{dt} = \sqrt{2V(x)} \Rightarrow \frac{d^2x}{dt^2} &= \frac{d}{dt}(\sqrt{2V(x)}),\\ &= \frac{dx}{dt}\frac{V'(x)}{\sqrt{2V(x)}},\\ &= V'(x).\end{aligned}$$

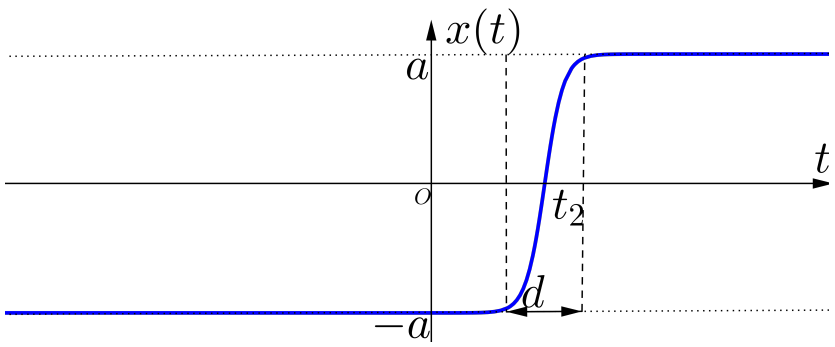


Figure 2.9: Diagram of $x(t)$ showing one instanton for the case $x_i = -a$ and $x_f = a$.

As mentioned above, the physical interpretation of the instanton and the anti-instanton are clear. The anti-instanton path reproduces the effect of a particle which is tunnelling through potential barrier from $x_i = a$ to $x_f = -a$. The instanton also corresponds to tunnelling through the potential barrier, but now from $x_i = -a$ to $x_f = a$. The transition occurs in a small time d so that the particle spends most of the time at rest at the minima of the potential.

The profile of the Euclidean Lagrangian L_E for the complete path comprising of an anti-instanton followed by an instanton is shown in Figure 2.10. The total area in this figure gives the total Euclidean action for the instanton and anti-instanton configuration.

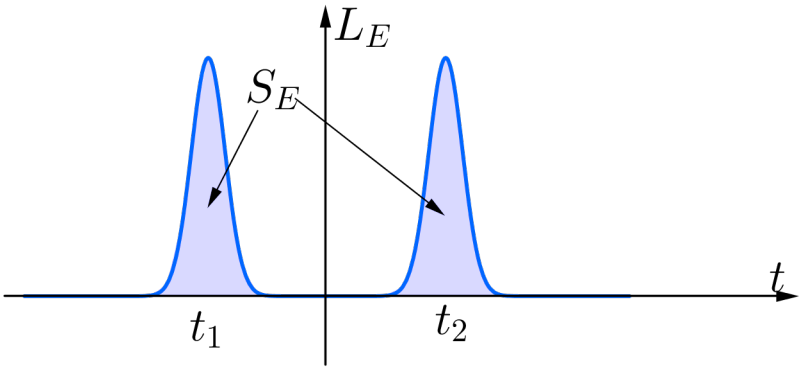


Figure 2.10: Profile of L_E for anti-instanton followed by instanton configuration

Since the action of the instanton and the anti-istanton are equal, the Euclidean action for an instanton plus an anti-instanton is double the Euclidean action for a single instanton (or single anti-instanton). If we denote the action for a single instanton by S_o, the total action in Figure 2.10 is

$$S_E = S_2 = 2S_o. \tag{2.4.10}$$

In general we can have many instantons and anti-instantons for the boundary condition $x_i = x_f = a$. The only condition is that the Euclidean motion should begin with an anti-instanton and end with an instanton and it is alternating sequence of instantons and anti-instantons. The total action for n pairs of instantons and anti-instantons is

$$S_n = 2nS_o. \tag{2.4.11}$$

For the three other boundary conditions the construction of instanton and anti-instanton configurations proceeds as for the case we have considered. The Euclidean action can start with either an instanton or

an anti-instanton and can terminate with either an instanton or an anti-instanton. The precise details in each case will be different because of the boundary conditions. Our analysis has assumed that our Euclidean motion is made up from a sequence of well separated instanton and anti-instanton motions. This is known as the dilute instanton gas approximation. The dilute instanton gas approximation can be invoked if

$$\frac{d}{|t_2 - t_1|} \ll 1 \tag{2.4.12}$$

is satisfied. In this last formula d is the size of the instanton (or anti-instanton) and $|t_2 - t_1|$ is the difference between time of adjacent instantons and anti-instantons in the Euclidean motion.

2.5 n-instantons

To compute the contribution from n instanton/anti-instanton configurations we need to compute

$$\left.\frac{N}{\sqrt{\det\left(-\partial_t^2 + V''\right)}}\right|_{n-\text{instantons}}. \tag{2.5.1}$$

It turns out that

$$\left.\frac{N}{\sqrt{\det\left(-\partial_t^2 + V''\right)}}\right|_{n-\text{instantons}} = \sqrt{\frac{2\omega}{\pi\hbar}} e^{-\omega T/2} K^n. \tag{2.5.2}$$

To determine the number K, we will compare (2.5.2) to our previous results. Notice that each instanton introduces a well in $V''(\tilde{x})$. We can think of K as the inverse square root of the product of the energies of any boundstates of the well times the effect of the well on the continuum eigenvalues. In Figure 2.11 we only illustrate the boundstate eigenvalues. This interpretation for K was developed in detail in [12].

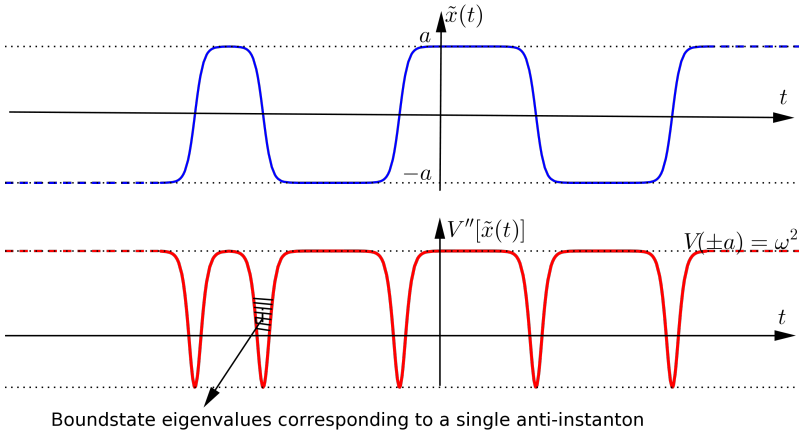


Figure 2.11: Configurations of intantons/anti-instantons and the corresponding profile of V''.

Now consider

$$\langle x_f|0\rangle\langle 0|x_i\rangle e^{-E_o T/\hbar} = \sum_n \int_{-T/2}^{T/2} dt_1 \int_{t_1}^{T/2} dt_2 \ldots \int_{t_{n-1}}^{T/2} dt_n e^{-nS_o/\hbar} \sqrt{\frac{2\omega}{\pi\hbar}} e^{-\frac{\omega T}{2}} K^n. \tag{2.5.3}$$

This expression sums all possible instanton configurations. For example, the first instanton centered at t_1 can be anywhere between $-T/2$ and $T/2$. The next instanton centered at t_2 should placed between t_1 and $T/2$. Repeating this process will give us the integrals in (2.5.3). Note that if we permute the integral bounded by t_i with those of t_j, $i \neq j$ final answer doesn't change. Generally, under the $n!$ possible permutations, the sum remains the same. We will now argue that

$$\int_{-T/2}^{T/2} dt_1 \int_{t_1}^{T/2} dt_2 \dots \int_{t_{n-1}}^{T/2} dt_n = \frac{T^n}{n!}. \tag{2.5.4}$$

The proof of (2.5.4) can be constructed by thinking of the left hand side as a fraction of the volume of an $n-$cube in n dimensions.

Proof. For $n = 1, 2$ and 3 respectively we have

$$\int_0^T dt_1 = \frac{T^1}{1!}, \tag{2.5.5}$$

$$\int_0^T dt_1 \int_{t_1}^T dt_2 = \int_0^T dt_1 \left[Tt_1 - \frac{t_1^2}{2} \right]_T^T = \frac{T^2}{2!}, \tag{2.5.6}$$

$$\int_0^T dt_1 \int_{t_1}^T dt_2 \int_{t_2}^T dt_3 = \int_0^T dt_1 \left(\frac{T^2}{2} - Tt_1 + \frac{t_1^2}{2} \right) = \frac{T^3}{3!}. \tag{2.5.7}$$

In fact we can think of (2.5.5), (2.5.6) and (2.5.7) as respectively the length of the segment a), the hatched surface of the square b) and the portion of the cube c) in the Figure 2.12. Note that for simplicity we shifted the integration region from $-T/2$ to $T/2$ to integration from 0 to T.

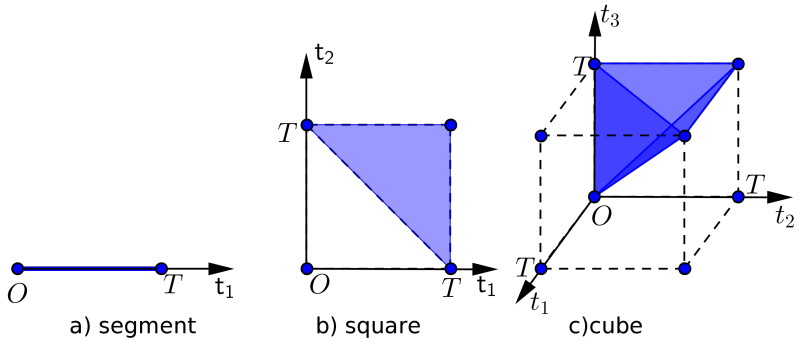


Figure 2.12: Figures of the first three $n-$cube volumes.

In general (2.5.4) is a portion of the n-cube with edge length T. The portion of the n-cube is delimited by $(n+1)$ vertices among the 2^n vertices of the $n-$cube. These $n+1$ vertices are diagonally linked as we can see above in Figure a), b) and c) for $n = 1, 2$ and 3. The total volume of the $n-$cube is T^n. There are $n!$ such portions inside the n-cube, which follows by counting the possibilities of joining $(n+1)$ vertices diagonally. In fact, our previous argument counting the number of possible permutations is related to the number of the portions of the volume that we consider. Finally a single portion of the $n-$cube has exactly the volume $v = \frac{T^n}{n!}$. □

As a consequence we can now rewrite (2.5.3) as

$$\langle x_f|0\rangle\langle 0|x_i\rangle e^{-E_oT/\hbar} = \sqrt{\frac{2\omega}{\pi\hbar}}e^{-\omega T/2}\sum_n \frac{(KTe^{-S_o/\hbar})^n}{n!}. \tag{2.5.8}$$

Having this expression, we can check the validity of the dilute instanton gas approximation by confirming that n is small. Consider

$$\begin{aligned}\sum_n \frac{X^n}{n!} &= \sum_n e^{n\log(X)-log(n!)},\\ &= \sum_n e^{n\log(X)-n\log n+n}.\end{aligned}$$

The summand is maximized when

$$\Rightarrow \frac{d}{dn}(n\log(X) - n\log n + n) = 0,$$

which implies

$$\Leftrightarrow X = n.$$

If we substitute X by $KTe^{-S_o/\hbar}$, this last equation implies the terms that dominate the sum have

$$n \leq 100X = 100KTe^{-S_o/\hbar},$$

which for $\hbar \to 0$ is small. Hence the dilute instanton gas approximation is valid because the sum is dominated by small n. Recall that we always require $\hbar \to 0$ to guarantee validity of the steepest method approximation for the path integral. In fact, all our expressions were derived in this limit.

For the boundary condition $x_i = x_f = \pm a$, n should be even because the motion always includes pairs of instantons and anti-instantons. And for the case $x_i = -x_f = \pm a$, n should be odd because the motion always includes either a single instanton or anti-instanton which is not paired. Furthermore, if n is even or odd, the summation factor in (2.5.8) can be written respectively as

$$\sum_n \frac{(KTe^{-S_o/\hbar})^n}{n!} = \frac{e^{KTe^{-S_o/\hbar}} + e^{-KTe^{-S_o/\hbar}}}{2} \quad \Leftrightarrow \quad n \text{ even}, \tag{2.5.9}$$

$$\sum_n \frac{(KTe^{-S_o/\hbar})^n}{n!} = \frac{e^{KTe^{-S_o/\hbar}} - e^{-KTe^{-S_o/\hbar}}}{2} \quad \Leftrightarrow \quad n \text{ odd}. \tag{2.5.10}$$

Thus, we have

$$\sum_{j=-}^{+}\langle \pm a|j\rangle\langle j|\pm a\rangle e^{-E_jT/\hbar} = \sqrt{\frac{2\omega}{\pi\hbar}}e^{-\omega T/2}\left[\frac{e^{KTe^{-S_o/\hbar}} + e^{-KTe^{-S_o/\hbar}}}{2}\right], \tag{2.5.11}$$

$$\sum_{j=-}^{+}\langle \pm a|j\rangle\langle j|\mp a\rangle e^{-E_jT/\hbar} = \sqrt{\frac{2\omega}{\pi\hbar}}e^{-\omega T/2}\left[\frac{e^{KTe^{-S_o/\hbar}} - e^{-KTe^{-S_o/\hbar}}}{2}\right]. \tag{2.5.12}$$

Finally the energy of the ground state can be read from (2.5.11) and (2.5.12) as

$$E_{\mp} = \frac{\hbar\omega}{2} \pm K\hbar e^{-S_o/\hbar}. \tag{2.5.13}$$

The associated wavefunctions obey

$$\langle a|+\rangle = \langle -a|+\rangle \quad \langle a|-\rangle = -\langle -a|-\rangle, \tag{2.5.14}$$

suggesting that they have a definite parity. The second term in the equation (2.5.13) is related to a single instanton because of the term S_o in the argument of the exponential. Note that it does not admit a power series expansion in $\hbar$. This is therefore a non-perturbative correction. We see that the ground state degeneracy is lifted by instantons, that is, because of quantum tunnelling. This simple example suggests that instantons will modify the vacuum structure of QFT. If we write

$$E_{\mp} = \sum_{p\geq 0} a_p \hbar^p \pm K\hbar e^{-S_o/\hbar},$$

it is clear that the non-perturbative correction gives a negligible contribution to the value of the energy. It does however give the total value of the splitting

$$\Delta E = E_+ - E_-.$$

Before continuing with the computation of K we need to explain a subtle point. Recall the equation of motion for the instanton

$$\frac{d^2x}{dt^2} = V'(x). \tag{2.5.15}$$

The equation of motion is invariant under time translation. Under time translation, the position of the instanton is shifted but its shape doesn't change. The equation of motion remains the same at $t+\epsilon$ and thus

$$\frac{d^2x(t+\epsilon)}{dt^2} = V'\big[x(t+\epsilon)\big]. \tag{2.5.16}$$

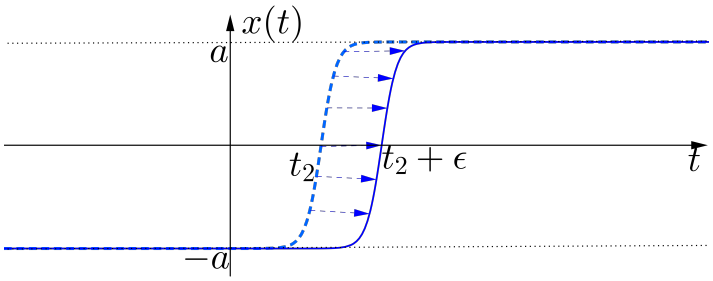


Figure 2.13: Profile of instanton under time translation.

After expanding (2.5.16) we recover the equation of motion at order $O(1)$

$$\frac{d^2x}{dt^2} = V'(x). \tag{2.5.17}$$

At order $O(\epsilon)$, we have

$$\frac{d^2}{dt^2}\left(\frac{dx}{dt}\right) = V''(x)\frac{dx}{dt}; \tag{2.5.18}$$

$$\Rightarrow \left(-\partial_t^2 + V''\big|_{\text{1-instanton}}\right)\left(\frac{dx}{dt}\right) = 0. \tag{2.5.19}$$

This last equation tells us that the eigenfunction $x_1 = \mathcal{N}\frac{dx}{dt}$ has the eigenvalue "0". The normalization constant $\mathcal{N}$ can be found by requiring that our eigenfunctions x_n are orthonormal. When $n = 1$, we have

$$\int_{-T/2}^{T/2} dt\big(x_1(t)\big)^2 = 1. \tag{2.5.20}$$

Hence, the normalization constant $\mathcal{N}$ should satisfy $\mathcal{N}^2 S_o = 1$, i.e.

$$x_1(t) = \frac{1}{\sqrt{S_o}}\frac{dx}{dt}. \tag{2.5.21}$$

x_1 is called a zero-mode because it corresponds to eigenvalue $\lambda = 0$. Note that (2.4.5) and (2.5.21) together yield

$$1 = \int_{-T/2}^{T/2} dt\big(x_1(t)\big)^2 = \int_{-T/2}^{T/2} dt\frac{1}{S_o}\left(\frac{dx}{dt}\right)^2; \tag{2.5.22}$$

$$S_o = \int_{-T/2}^{T/2} dt\left(\frac{1}{2}\left(\frac{dx}{dt}\right)^2 + V(x)\right). \tag{2.5.23}$$

The last equation is nothing but the definition of S_o, which is the Euclidean action for a single instanton (or anti-instanton). Using the equality of $\frac{dx}{dt}$ and $\sqrt{2V}$, we can also write

$$S_o = \int_{-a}^{a} dx\sqrt{2V(x)}. \tag{2.5.24}$$

2.5.1 Remark. In the previous calculations we have used time translation symmetry and, as a consequence of invariance of the system, we obtained the zero-mode $x_1(t)$. In general it is straightforward to see that there should be a zero-mode associated to each symmetry.

Recall that for a general path, we had

$$x(t) = \tilde{x}(t) + C_1 x_1(t) + \sum_{n\geq 2} C_n x_n(t); \text{ so that } dx(t) = dC_1 x_1(t) + \dots .$$

If $x(t)$ is a solution for 1-instanton it should satisfy (2.5.21), so that

$$x(t + dt_1) = x(t) + dt_1\frac{dx}{dt}.$$

Comparing the previous two equations gives

$$dC_1 = \sqrt{S_o}dt_1, \Rightarrow \frac{dC_1}{\sqrt{2\pi\hbar}} = \sqrt{\frac{S_o}{2\pi\hbar}}dt_1. \tag{2.5.25}$$

As a consequence we have

$$\frac{Ne^{-S[\tilde{x}]/\hbar}}{\sqrt{\det(-\partial_t^2 + V''[\tilde{x}])}} = \sqrt{\frac{S_o}{2\pi\hbar}}\frac{Ne^{-S[\tilde{x}]/\hbar}}{\sqrt{\det'(-\partial_t^2 + V''[\tilde{x}])}}. \tag{2.5.26}$$

The $\det(\dots)$ becomes $\det'(\dots)$ because we have to exclude the zero mode. Indeed, the integration over the zero mode was already accounted for when we integrated over the center of the instanton. However,

the measure for integration over the center of the instanton and integration over the zero mode are related as in (2.5.25) so to keep the whole expression consistent we included the factor $\sqrt{\frac{S_o}{2\pi\hbar}}$ in the expression with $\det'(\ldots)$.

Note that we only have time translation invariance (and hence the zero mode) in the limit $T \to \infty$. If we study the system for finite T, we don't have time translation invariance for our system and λ_1 doesn't vanish.

2.6 Computation of K

To avoid the zero-mode, we will work with a finite time T. We will compute $\det(-\partial_t^2 + V''[\tilde{x}])$. We are not going to redo all of the details of this calculation. The computation we perform now is a minor change of the computation of section 2.3, reflecting the fact that the potential of the system is no longer that of the harmonic oscillator. Recall that

$$\langle a|e^{-HT/\hbar}|-a\rangle_{1-\text{instanton}} = \frac{NTe^{-S_o/\hbar}}{\sqrt{\det'(-\partial_t^2 + V''|_{1-\text{instanton}})}}\sqrt{\frac{S_o}{2\pi\hbar}}, \tag{2.6.1}$$

$$= \frac{NTe^{-S_o/\hbar}}{\sqrt{\det(-\partial_t^2+\omega^2)}}K; \tag{2.6.2}$$

$$\Rightarrow K = \sqrt{\frac{S_o}{2\pi\hbar}}\sqrt{\frac{\det(-\partial_t^2+\omega^2)}{\det'(-\partial_t^2 + V''|_{1-\text{insanton}})}}. \tag{2.6.3}$$

The determinant is expressed in terms of the function

$$(-\partial_t^2 + V'')\psi_0(t) = 0, \tag{2.6.4}$$

which is exactly what we had in our discussion of the harmonic oscillator. In addition, we already have a solution of this equation, in (2.5.19). Hence, we identify

$$\psi_0(t) = \frac{1}{\sqrt{S_o}}\frac{dx}{dt}. \tag{2.6.5}$$

For one instanton we have

$$(-\partial_t^2+\omega^2)\psi_0(t) = 0 \Rightarrow \psi_0(t) = \frac{1}{\sqrt{S_o}}\frac{d\tilde{x}}{dt} \to Ae^{-\omega t}, \quad \text{in the large } t \text{ limit.} \tag{2.6.6}$$

We can also check that if y_1 is a second solution we have

$$\frac{d}{dt}\left(x_1\frac{dy_1}{dt} - y_1\frac{dx_1}{dt}\right) = 0, \tag{2.6.7}$$

$$\Leftrightarrow \left(x_1\frac{dy_1}{dt} - y_1\frac{dx_1}{dt}\right) = C \quad \text{(constant)}. \tag{2.6.8}$$

Knowing the form of $x_1(t)$ (2.6.8) will allow us to determine the form of $y_1(t)$. (2.6.8) is known as the Wronskian of the two solutions. Choose the constant C

$$C = 2A\omega, \tag{2.6.9}$$

so that $y_1(t)$ is

$$y_1(t) = Ae^{\omega t}.$$

Note that this expression for $y_1(t)$ is only valid for large values of t because the expression we used for $x_1(t)$, assumes large t. Having the two solutions $x_1(t)$ and $y_1(t)$, the general solution to (2.6.4) is

$$\psi_0(t) = b_1 x_1(t) + b_2 y_2(t), \tag{2.6.10}$$

where $\{b_i\}_{i=1,2}$ are real constants. Use the boundary conditions to fix the values of b_i. The boundary conditions that $\psi_0(t)$ satisfies are

$$\psi_0(-T/2) = 0 \quad \text{and} \quad \left.\frac{d\psi_0(t)}{dt}\right|_{t=-T/2} = 1. \tag{2.6.11}$$

Substituting (2.6.10) into (2.6.11) leads to

$$b_1 = \frac{1}{2A\omega} e^{\omega T/2} \qquad b_2 = \frac{1}{2A\omega} e^{-\omega T/2}.$$

Our final solution is now

$$\psi_0(t) = \frac{1}{2A\omega}(e^{\omega T/2} x_1(t) + e^{-\omega T/2} y_1(t)). \tag{2.6.12}$$

This last expression for $\psi_0(t)$ yields

$$\psi_0(T/2) = \frac{1}{\omega}. \tag{2.6.13}$$

Note that we have

$$\frac{\det(-\partial_t^2 + \omega^2)}{\lambda_0 \det'(-\partial_t^2 + V'')} = \frac{\det(-\partial_t^2 + \omega^2)}{\det(-\partial_t^2 + V'')} \tag{2.6.14}$$

In the case of the simple harmonic oscillator we know that

$$\psi_{h.o}(t) = \frac{1}{\omega} \sinh\left((\frac{T}{2} + t)\omega\right); \tag{2.6.15}$$

$$\Rightarrow \psi_{h.o}(\frac{T}{2}) \underset{T\to\infty}{=} \frac{1}{2\omega} e^{\omega T}. \tag{2.6.16}$$

Thus by use of (2.6.13), (2.6.14) and (2.6.16) the expression for K in (2.6.3) becomes

$$K = \sqrt{\frac{S_o}{2\pi\hbar}} e^{\omega T/2} \sqrt{\frac{\lambda_0}{2}}. \tag{2.6.17}$$

To determine λ_0 solve

$$(-\partial_t^2 + V'')\psi_\lambda(t) = \lambda\psi_\lambda(t), \tag{2.6.18}$$

using the Green's function $G(t, t')$ [13]. As usual, we write

$$\psi_\lambda(t) = \psi_0(t) + \lambda \int_{-T/2}^{T/2} dt' G(t, t')\psi_\lambda(t'). \tag{2.6.19}$$

The Green's function obeys

$$(-\partial_t^2 + V'')G(t, t') = \delta(t - t'); \tag{2.6.20}$$

where $\delta(t-t')$ is the Dirac delta distribution. In addition, $\psi_0(t)$ obeys

$$(-\partial_t^2 + V'')\psi_0(t) = 0.$$

The boundary conditions for our problem are

$$\psi_0(-T/2) = 0 \quad \text{and} \quad \left.\frac{d\psi_0(t)}{dt}\right|_{t=-T/2} = 1.$$

The second condition above again determines the otherwise arbitrary normalization of our solution. Now if we define $\theta(t-t')$, to be the Heaviside step function

$$\theta(t-t') = \begin{cases} 0 & \text{if } t' \geq t \\ 1 & \text{if } t' < t \end{cases},$$

the Green's function is given by

$$G(t,t') = \mathcal{N}\left[\theta(t-t')y_1(t)x_1(t') + \theta(t'-t)x_1(t)y_1(t')\right]. \tag{2.6.21}$$

Integrating the equation (2.6.20), using the Green's function (2.6.21), leads us to

$$\int_{t'-\epsilon}^{t'+\epsilon} dt(-\partial_t^2 + V'')G(t,t') = \int_{t'-\epsilon}^{t'+\epsilon} dt\delta(t-t') = 1; \tag{2.6.22}$$

$$= -\left.\frac{dG}{dt}\right|_{t'-\epsilon}^{t'+\epsilon} = -\mathcal{N}[x_1\dot{y}_1 - \dot{x}_1 y_1] = -2\mathcal{N}\omega A = 1, \tag{2.6.23}$$

$$\Rightarrow \mathcal{N} = -\frac{1}{2\omega A}. \tag{2.6.24}$$

Note that we have used (2.6.9) to obtain (2.6.23). Now, returning to (2.6.21), and using the above results, we find

$$G(t,t') = -\frac{1}{2\omega A}\left[\theta(t-t')y_1(t)x_1(t') + \theta(t'-t)x_1(t)y_1(t')\right]. \tag{2.6.25}$$

Thus, for eigenvalue λ_0 we have

$$\psi_{\lambda_0}(t) = \psi_0(t) + \lambda_0 \int_{-T/2}^{T/2} dt' G(t,t')\psi_{\lambda_0}(t'). \tag{2.6.26}$$

Evaluating at $T/2$ we find

$$\psi_{\lambda_0}(T/2) = 0 = \frac{1}{\omega} + \lambda_0 \int_{-T/2}^{T/2} dt' G(t,t')\psi_0(t'), \tag{2.6.27}$$

$$= \frac{1}{\omega} + \lambda_0\left(-\frac{1}{2\omega A}\right)\int_{-T/2}^{T/2} dt' y_1(T/2)x_1(t')\left[\frac{1}{2A\omega}\left(e^{\omega T/2}x_1(t') + e^{-\omega T/2}y_1(t')\right)\right] \quad (t > t'), \tag{2.6.28}$$

$$= \frac{1}{\omega} + \lambda_0\left(\frac{-e^{\omega T}}{4A^2\omega^2}\right)\int_{-T/2}^{T/2} dt' x_1^2(t'), \tag{2.6.29}$$

$$= \frac{1}{\omega} + \lambda_0 e^{\omega T}\left(-\frac{1}{4A^2\omega^2}\right); \tag{2.6.30}$$

$$\Leftrightarrow \lambda_0 = 4\omega A^2 e^{-\omega T} \Leftrightarrow \lambda_0 \to 0, \quad \text{when } T \to \infty. \tag{2.6.31}$$

The passage to the third and fourth lines has used the orthogonality of x_1 and y_1. At this stage if we plug this last expression into (2.6.17) we obtain the value

$$K = \sqrt{\frac{S_o}{\pi\hbar\omega}}A. \tag{2.6.32}$$

This completes the computation of ground state energies using instantons.

2.7 A concrete computation

In this section we will consider a specific potential and carry out the explicit calculation of the number K. We have mentioned previously that the number K can be interpreted as the inverse square root of the product of the energies of any boundstates of a well in V'' times the zero mode measure and a factor measuring the effect of the well on the continuum eigenvalues. Here we demonstrate this statement by direct calculation. We study the following potential

$$V(x) = gx^4 - x^2 + \frac{1}{4g}, \tag{2.7.1}$$

where g is positive real parameter. Clearly this potential is an example of a double well potential. Note that the potential $V(x)$ can be written as

$$V(x) = g\left(x^2 - \frac{1}{2g}\right)^2. \tag{2.7.2}$$

Consider the equation for the instanton obtained in (2.4.9),

$$\frac{dx}{dt} = \pm\sqrt{2V(x)}. \tag{2.7.3}$$

Inserting the expression for V in (2.7.2) into this equation yields

$$\frac{dx}{dt} = \pm\left(\frac{(\sqrt{2g}x)^2 - 1}{\sqrt{2g}}\right). \tag{2.7.4}$$

Now integrate this equation to get a solution $\tilde{x}(t)$, which is an extremum of the Euclidean action. We find

$$\pm\frac{\sqrt{2g}dx}{(\sqrt{2g}x)^2 - 1} = dt, \tag{2.7.5}$$

$$\pm\int_0^{\tilde{x}} \frac{\sqrt{2g}dx'}{(\sqrt{2g}x')^2 - 1} = \int_{t_2}^{t} dt = (t - t_2). \tag{2.7.6}$$

The admissible boundary conditions are $x_i = \frac{\pm 1}{\sqrt{2g}}$, $x_f = \frac{\pm 1}{\sqrt{2g}}$. Furthermore, for our Euclidean motions we always have

$$\frac{-1}{\sqrt{2g}} < \tilde{x}(t) < \frac{1}{\sqrt{2g}}. \tag{2.7.7}$$

Hence,

$$\pm\int_0^{\sqrt{2g}\tilde{x}} \frac{dx'}{1 - (x')^2} = (t - t_2), \tag{2.7.8}$$

$$\pm\text{atanh}(\sqrt{2g}\tilde{x}) = (t - t_2), \tag{2.7.9}$$

$$\Rightarrow \tilde{x} = \pm\frac{\tanh(t - t_2)}{\sqrt{2g}}. \tag{2.7.10}$$

The profile of $\tilde{x}(t)$ matches what we sketched in Figure 2.9. Having the above expression for $\tilde{x}$, we can now compute $V''[\tilde{x}]$. We find

$$V''(x) = 12gx^2 - 2, \tag{2.7.11}$$

$$\Rightarrow V''(\tilde{x}) = 6\tanh^2(t - t_2) - 2. \tag{2.7.12}$$

We also need to evaluate S_o since the expression for K in (2.6.32) depends on it. Recall that

$$S_o = \int_{\frac{-1}{\sqrt{2g}}}^{\frac{1}{\sqrt{2g}}} dx\sqrt{2V(x)}. \tag{2.7.13}$$

Inserting the expression for $V(x)$ obtained in (2.7.2), we find

$$S_o = \int_{\frac{-1}{\sqrt{2g}}}^{\frac{1}{\sqrt{2g}}} dx\sqrt{2g\left(\frac{1}{2g} - x^2\right)^2}, \tag{2.7.14}$$

$$= 2\sqrt{2g}\int_0^{\frac{-1}{\sqrt{2g}}}\left(\frac{1}{2g} - x^2\right), \tag{2.7.15}$$

$$= 2\sqrt{2g}\left(\frac{x}{2g}\Big|_0^{1/\sqrt{2g}} - \frac{x^3}{3}\Big|_0^{1/\sqrt{2g}}\right), \tag{2.7.16}$$

$$\Rightarrow S_o = \frac{2}{3g}. \tag{2.7.17}$$

A quick calculation shows

$$\omega^2 = V''(\pm 1/\sqrt{2g}) = 4 = 2^2; \Rightarrow \omega = 2. \tag{2.7.18}$$

To find the value of A, we proceed as follows. From (2.6.6) we know that for large t we have

$$\frac{1}{\sqrt{S_o}}\frac{d\tilde{x}}{dt} = Ae^{-\omega t}. \tag{2.7.19}$$

Substituting the values of $\tilde{x}(t)$, ω and S_o we find

$$Ae^{-2t} = \frac{1}{\sqrt{\frac{2}{3g}}}\frac{1}{\sqrt{2g}}\frac{1}{\cosh^2(t)} = \frac{\sqrt{3}}{2}\frac{1}{\cosh^2(t)} \tag{2.7.20}$$

For large t, $\cosh^2(t)$ is approximatively e^{2t}. Hence, the value of A in this limit is

$$Ae^{-2t} = \frac{\sqrt{3}}{2}e^{-2t}; \tag{2.7.21}$$

$$\Rightarrow A = \frac{\sqrt{3}}{2}. \tag{2.7.22}$$

Now, substitute the above values into the expression for K found in (2.6.32) to obtain

$$K = \sqrt{\frac{S_o}{\pi\hbar\omega}}A = \sqrt{\frac{\frac{2}{3g}}{2\pi\hbar}}\frac{\sqrt{3}}{2}. \tag{2.7.23}$$

Finally, the value of K is

$$K = \sqrt{\frac{1}{\pi g\hbar}}. \tag{2.7.24}$$

We want to show that K is equal to the product of the zero mode measure $\sqrt{\frac{S_o}{2\pi\hbar}}$ with the inverse square root of the product of the energies of any boundstates of a well in V'' and a factor $\sqrt{\Phi^{-1}}$ measuring the effect of the well on the continuum eigenvalues of the following equation

$$\left(-\frac{d^2}{dt^2}+V''(\tilde{x})\right)\psi_\lambda=\lambda\psi_\lambda \tag{2.7.25}$$

with the boundary condition

$$\psi_\lambda(\pm T/2)=0.$$

Concretely we want to prove that

$$K=\sqrt{\frac{1}{\pi g\hbar}}=\sqrt{\prod_i \lambda_i^{-1}}\sqrt{\Phi^{-1}}\sqrt{\frac{S_o}{2\pi\hbar}} \tag{2.7.26}$$

where i runs over the boundstates and Φ is the total contribution from the continuum eigenvalues of (2.7.25). In this section we will not compute Φ because a very detailed and complete calculation of Φ can be found at [12, p. 220-222], thus we just give the result found in this source $\Phi=\frac{1}{9}$. For our problem (2.7.25) is given by

$$\left(-\frac{d^2}{dt^2}+6\tanh^2(t-t_2)-2\right)\psi_\lambda=\lambda\psi_\lambda. \tag{2.7.27}$$

Our discussion makes use of [14, p. 72-73]. Choosing $t_2=0$, (2.7.27) can be written as

$$\frac{d^2}{dt^2}\Psi+\left(\lambda+2-6\tanh^2(t)\right)\Psi=0. \tag{2.7.28}$$

Use the new variable

$$\xi=\tanh(t).$$

Using

$$\frac{d\xi}{dt}=1-\tanh^2(t)=1-\xi^2, \tag{2.7.29}$$

$$\frac{d^2\xi}{dt^2}=\frac{d}{dt}[1-\tanh^2(t)]=-2[1-\tanh^2(t)]\tanh(t)=-2(1-\xi^2)\xi, \tag{2.7.30}$$

and the chain rule property of the differentiation, we have

$$\frac{d^2\Psi}{dt^2}=\frac{d}{dt}\frac{d\Psi}{dt}=\frac{d}{dt}\left(\frac{d\xi}{dt}\frac{d\Psi}{d\xi}\right), \tag{2.7.31}$$

$$=\frac{d^2\xi}{dt^2}\frac{d\Psi}{d\xi}+\left(\frac{d\xi}{dt}\right)^2\frac{d^2\Psi}{d\xi^2}, \tag{2.7.32}$$

$$=-2(1-\xi^2)\xi\frac{d\Psi}{d\xi}+(1-\xi^2)^2\frac{d^2\Psi}{d\xi^2}. \tag{2.7.33}$$

(2.7.28) now becomes

$$-2(1-\xi^2)\xi\frac{d\Psi}{d\xi}+(1-\xi^2)^2\frac{d^2\Psi}{d\xi^2}+[\lambda-4+6(1-\xi^2)]\Psi=0, \tag{2.7.34}$$

which can be written as

$$\frac{d}{d\xi}\left((1-\xi^2)\frac{d\Psi}{d\xi}\right)+\left(2(2+1)-\frac{4-\lambda}{1-\xi^2}\right)\Psi=0. \tag{2.7.35}$$

It turns out that (2.7.35) is solved by associated Legendre polynomials of degree $l=2$ and order $m=\sqrt{4-\lambda}$. In fact, the associated Legendre polynomials of degree l and order m are defined by

$$\frac{d}{dx}\left((1-x^2)\frac{d\Phi_l^m(x)}{dx}\right)+\left(l(l+1)-\frac{m^2}{1-x^2}\right)\Phi_l^m(x)=0, \tag{2.7.36}$$

where l and m are integers and $x\in[-1,\ 1]\subset\mathbb{R}$. In addition, only pairs of integers l and m satisfying $0\leq m\leq l$ yield a unique non trivial solution defined by

$$\Phi_l^m(x)=(-1)^m(1-x^2)^{\frac{m}{2}}\frac{d^{l+m}}{dx^{l+m}}(x^2-1)^l. \tag{2.7.37}$$

Hence (2.7.35) has two normalizable solutions for $l=2$ if and only if

$$m=\sqrt{4-\lambda}\in\mathbb{N}\cup\{0\}\Leftrightarrow\lambda=\begin{cases}\lambda_o=0 & \Rightarrow m=2\leq l=2,\\ \lambda_1=3 & \Rightarrow m=1<l=2.\end{cases} \tag{2.7.38}$$

Note that $\lambda=4$ also yields a Legendre polynomial, but its associated eigenfunction $\Phi_2^0(\xi)$ is not normalizable. Thus, (2.7.25) has exactly two eigenvalues

$$\lambda_o=0, \tag{2.7.39}$$
$$\lambda_1=3. \tag{2.7.40}$$

Another way of finding these values is to numerically solve the partial differential equation (2.7.25). In the appendix Chapter A.1 we carry out this exercise and show that the only normalizable eigenfunctions for (2.7.25) indeed correspond to the above eigenvalues. $\lambda_o=0$ is the eigenvalue associated to the zero mode and we should have expected its presence. Now, we have

$$K=\sqrt{\frac{S_o}{2\pi\hbar}}\frac{1}{\sqrt{3}}\sqrt{\Phi^{-1}}=\sqrt{\frac{\frac{2}{3g}}{2\pi\hbar}}\frac{1}{\sqrt{3}}\sqrt{9}, \tag{2.7.41}$$
$$=\sqrt{\frac{1}{\pi g\hbar}}. \tag{2.7.42}$$

Hence, there is complete agreement between (2.7.24) and (2.7.42). This shows that K is indeed the product of the zero mode measure with the inverse square root of the product of the energies of any boundstates of a well in V'' times the effect of the well on the continuum eigenvalues of

$$\left(-\partial_t^2+V''\right)\psi_{\lambda_i}(t)=\lambda_i\psi_{\lambda_i}(t). \tag{2.7.43}$$

For a single instanton, i.e. a single dip of the potential V'', the potential is plotted in the following figure.

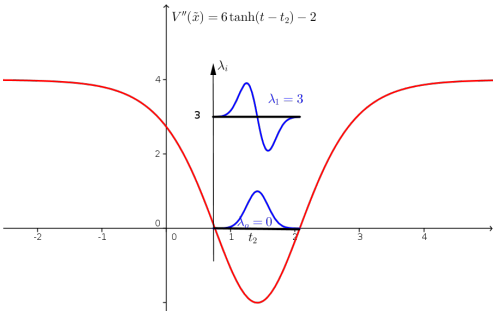


Figure 2.14: Boundstate eigenvalues of the potential V''.

2.8 Band energy and instantons

In this section our discussion follows [15]. Our starting point was the simple harmonic oscillator. In this case the Euclidean action has only one extremum. Next we considered a double well potential. Since this potential has two minima, we showed that the Euclidean action has an infinite number of extrema. These extrema were a succession of instantons and anti-instantons chosen to be consistent with the boundary conditions. We are now going to consider a potential that is a one dimensional periodic function of position. In this case any collection of instantons and/or anti-instantons gives an extremum of the Euclidean action. We consider the potential

$$V(x) = \frac{1}{2} - \frac{\cos(2\pi x)}{2}.$$

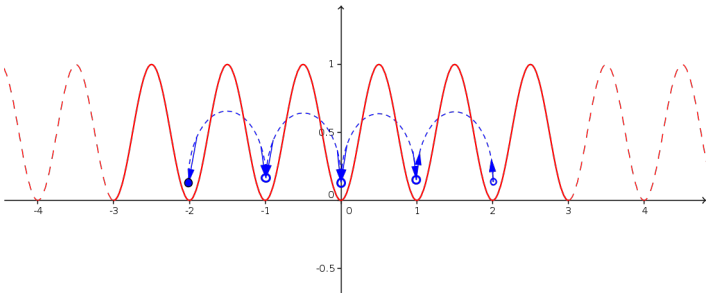


Figure 2.15: Profile of $V(x)$ together with a particle tunnelling through the barriers of potential from minima to minima

As seen in Figure 2.15, the allowed tunnelling transitions has dramatically increased. For example the number of instantons can considerably exceed the number of anti-instantons. To see what is going on we draw an example of a configuration of instantons and anti-instantons in Figure 2.16.

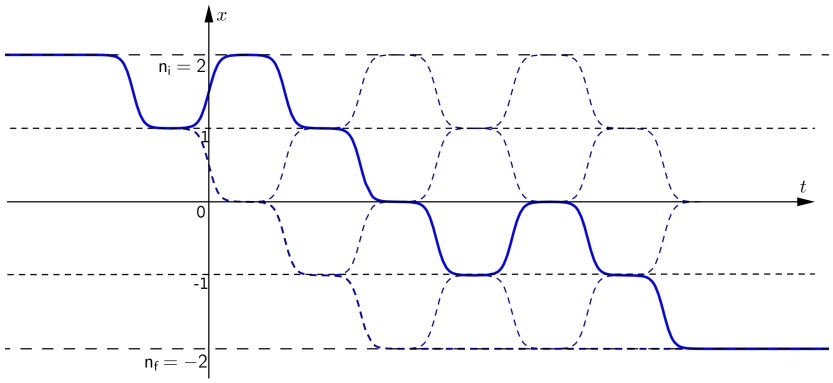


Figure 2.16: Example of a configuration for particle tunnelling from $x = 2$ to $x = -2$

For simplicity denote

$$\begin{cases} \tilde{n} & \text{numbers of anti-instantons} \\ n & \text{numbers of instantons.} \end{cases}$$

We can easily see from Figure 2.16 that n and $\tilde{n}$ satisfy

$$n_f - n_i = \tilde{n} - n, \tag{2.8.1}$$

where n_i and n_f are respectively the initial and final position of the particle. As usual, we want to compute

$$\langle n_f|0\rangle\langle 0|n_i\rangle e^{-E_o T/\hbar}.$$

In our semiclassical approximation we have

$$\langle n_f|0\rangle\langle 0|n_i\rangle e^{-E_o T/\hbar} = \sum_{n\geq 0}\sum_{\tilde{n}\geq 0} \delta_{n_f - n_i + n - \tilde{n}} e^{-(n+\tilde{n})\frac{S_o}{\hbar}} \left(\sqrt{\frac{2\omega}{\pi\hbar}} e^{-\omega T/2} K^{n+\tilde{n}}\right) \frac{T^{n+\tilde{n}}}{n!\tilde{n}!}. \tag{2.8.2}$$

Now, note that

$$\delta_{n_f - n_i + n - \tilde{n}} = \int_0^{2\pi} d\theta e^{i2\pi(n_f - n_i + n - \tilde{n})\theta}. \tag{2.8.3}$$

Thus (2.8.2) can be rewritten as two unconstrained sums

$$\langle n_f|0\rangle\langle 0|n_i\rangle e^{-E_oT/\hbar} = \sum_{n\geq 0}\sum_{\tilde{n}\geq 0}\int_0^{2\pi} d\theta e^{i2\pi(n_f-n_i+n-\tilde{n})\theta} e^{-(n+\tilde{n})\frac{S_o}{\hbar}} \left(\sqrt{\frac{2\omega}{\pi\hbar}}e^{-\omega T/2}K^{n+\tilde{n}}\right)\frac{T^{n+\tilde{n}}}{n!\tilde{n}!}, \tag{2.8.4}$$

$$= \int_0^{2\pi} d\theta\sqrt{\frac{2\omega}{\pi\hbar}}e^{-\frac{\omega T}{2}+i2\pi(n_f-n_i)\theta}\sum_{n\geq 0,\tilde{n}\geq 0}\frac{\left(KTe^{-\frac{S_o}{\hbar}}e^{i2\pi\theta}\right)^n}{n!}\frac{\left(KTe^{-\frac{S_o}{\hbar}}e^{-i2\pi\theta}\right)^{\tilde{n}}}{\tilde{n}!}, \tag{2.8.5}$$

$$= \int_0^{2\pi} d\theta\sqrt{\frac{2\omega}{\pi\hbar}}e^{-\frac{\omega T}{2}+i2\pi(n_f-n_i)\theta}\exp\left(KTe^{-\frac{S_o}{\hbar}}e^{i2\pi\theta}\right)\exp\left(KTe^{-\frac{S_o}{\hbar}}e^{-i2\pi\theta}\right), \tag{2.8.6}$$

$$= \int_0^{2\pi} d\theta\sqrt{\frac{2\omega}{\pi\hbar}}e^{-\frac{\omega T}{2}+i2\pi(n_f-n_i)\theta}\exp\left(2KTe^{-\frac{S_o}{\hbar}}\cos(2\pi\theta)\right), \tag{2.8.7}$$

$$= \int_0^{2\pi} d\theta\sqrt{\frac{2\omega}{\pi\hbar}}e^{i2\pi(n_f-n_i)\theta}e^{[-\frac{\hbar\omega}{2}+2\hbar Ke^{-\frac{S_o}{\hbar}}\cos(2\pi\theta)]\frac{T}{\hbar}} \tag{2.8.8}$$

Clearly, (2.8.8) permits us to identify the energy

$$E_o = \frac{\hbar\omega}{2} - 2\hbar Ke^{-S_o/\hbar}\cos(2\pi\theta), \tag{2.8.9}$$

and the wavefunction

$$\langle n_f|0\rangle\langle 0|n_i\rangle = \sqrt{\frac{2\omega}{\pi\hbar}}e^{i2\pi(n_f-n_i)\theta}. \tag{2.8.10}$$

So the single infinitely degenerate ground state has been replaced by a continuous energy spectrum, parametrized by θ. These results are familiar from solid-state physics. The degenerate ground state energy has been replaced by a band of allowed energy levels. In solid state physics these bands and (and the gaps between bands, called band gaps) are derived by examining the allowed quantum mechanical wavefunctions for an electron in a large, periodic lattice of atoms or molecules. Band theory explains many physical properties of solids, including electrical resistivity and optical absorption. Here we have obtained the lowest band of allowed energies by summing over instanton and anti-instanton configurations.

2.9 Schrödinger equation description of the double well potential

In this section we want to recover the expression for the ground state energy found in (2.5.13)

$$E_\mp = \frac{\hbar}{2} \pm \sqrt{\frac{\hbar S_o}{\pi}}Ae^{-S_o/\hbar}. \tag{2.9.1}$$

Note that above we set $\omega = 1$ and use the expression for K found in (2.6.32). To recover this result we are going to solve the Schrödinger equation using the **WKB** approximation. Our motivation is to verify that the instanton calculation is in perfect agreement with standard quantum mechanics. The time independent Schrödinger equation in one dimension is

$$H\Psi(x) = \left(\frac{-\hbar^2}{2m}\frac{d^2}{dx^2} + V(x)\right)\Psi(x) = E\Psi(x). \tag{2.9.2}$$

When the potential $V(x)$ is constant, $\Psi(x)$ is a plane wave. It is straightforward to see that the plane wave solution, in the case of a constant potential, has the form

$$\Psi(x) = \Psi_o e^{\frac{i}{\hbar}px},\ p = \pm\sqrt{2m(E-V)}.$$

Of course, p is constant because we assumed that V is constant. Lets continue to use this formula, even when we have a position dependent potential. We write

$$\Psi(x) = e^{\frac{i}{\hbar}\int_0^x dyp(y)} \text{ where } (p(y))^2 = 2m[E-V(y)], \tag{2.9.3}$$

This is not a solution. However, if we set

$$\Psi(x) = e^{\frac{i}{\hbar}\int_0^x dyp(y)+\sum_{i\geq 1}\hbar^{i-1}\phi_i(x)}, \tag{2.9.4}$$

and plug it into the Schrödinger equation,

$$\left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}+V(x)\right)e^{\frac{i}{\hbar}\int_0^x dyp(y)+i\phi_1(x)+i\hbar\phi_2(x)+O(\hbar^2)} = Ee^{\frac{i}{\hbar}\int_0^x dyp(y)+i\phi_1(x)+i\hbar\phi_2(x)+O(\hbar^2)} \tag{2.9.5}$$

we have

$$\frac{-\hbar^2}{2m}\frac{d}{dx}\Psi(x) = \frac{-\hbar^2}{2m}\left(\frac{i}{\hbar}p(x)+i\phi_1'(x)+O(\hbar)\right)\Psi(x), \tag{2.9.6}$$

$$\Rightarrow \frac{-\hbar^2}{2m}\frac{d^2}{dx^2}\Psi(x) = \frac{-\hbar^2}{2m}\left(\frac{i}{\hbar}p'(x)+i\phi_1''(x)+O(\hbar)\right)\Psi(x) - \frac{\hbar^2}{2m}\left(\frac{i}{\hbar}p(x)+i\phi_1'(x)+O(\hbar)\right)^2\Psi(x). \tag{2.9.7}$$

Hence we get

$$\text{for } O(\hbar^0): \quad \frac{p^2(x)}{2m}+V(x) = E, \tag{2.9.8}$$

$$\text{for } O(\hbar^1): \quad \frac{-i\hbar}{2m}p'(x)+\frac{\hbar}{m}p(x)\phi_1'(x) = 0. \tag{2.9.9}$$

Clearly at first order in $\hbar$, which is given in equation (2.9.8) we recover the definition of $p(x)$ itself. At the second order given in equation (2.9.9) we can obtain $\phi_1(x)$ by solving this an ordinary differential equation. We find

$$\frac{-i}{2}p'(x)+p(x)\phi_1'(x) = 0 \Leftrightarrow \phi_1'(x) = \frac{i}{2}\frac{p'(x)}{p(x)}, \tag{2.9.10}$$

$$\Rightarrow \phi_1(x) = \frac{d}{dx}\left(i\log\left(\sqrt{p(x)}\right)\right). \tag{2.9.11}$$

Finally, by choosing the constant of integration to be zero, the WKB wavefunction is

$$\Psi_{\text{WKB}} = \frac{e^{\frac{i}{\hbar}\int_0^x dyp(y)}}{\sqrt{p(x)}}. \tag{2.9.12}$$

We need to know where the WKB wavefunction provides a valid description. Ψ_{WKB} is only valid away from the turning points of the classical motion. Since we are considering particles of very small energy, the turning points of the classical motion are near the bottom of each well. To solve our problem, we

will use the WKB approximation far from the bottom of each well and the harmonic oscillator solution near the bottom of each well. To obtain the complete wavefunction we will match these two descriptions in the region in which they overlap. This is illustrated in Figure 2.17.

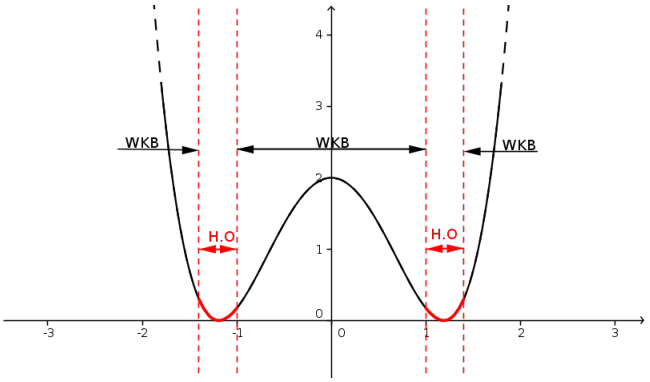


Figure 2.17: Figure illustrating our description.

Thus far from the wells, we transform $p(x) \to ik(x)$ so that our WKB wavefunction becomes

$$\Psi_{\text{WKB}}(x) = \frac{1}{\sqrt{k(x)}} e^{\frac{-1}{\hbar}\int_0^x dy k(y)}, \; k(y) = \sqrt{2m(E - V(y))}. \tag{2.9.13}$$

Far from the minima of the potential

$$E \ll V(x).$$

In this case,

$$k(x) = \sqrt{2(V(x) - E)}, \tag{2.9.14}$$

$$= \sqrt{2V(x)}\sqrt{1 - \frac{E}{V(x)}} = \sqrt{2V(x)}\left(1 - \frac{E}{2V(x)}\right) + \cdots, \tag{2.9.15}$$

$$= \sqrt{2V(x)} - \frac{E}{\sqrt{2V(x)}} + \cdots. \tag{2.9.16}$$

Integrating the above expression yields

$$\int_0^x dy k(y) = \int_0^x dy\sqrt{2V(y)} - \int_0^x dy \frac{E}{\sqrt{2V(y)}} + \cdots. \tag{2.9.17}$$

It turns out that we can relate the integral of $k(x)$ to the Euclidean action S_o of a single instanton. In fact, the first term can be rewritten as

$$\int_0^x dy\sqrt{2V(y)} = \int_0^a dy\sqrt{2V(y)} + \int_a^x dy\sqrt{2V(y)}. \tag{2.9.18}$$

Since $V(x)$ is an even function, we have

$$\int_0^a dy\sqrt{2V(y)} = \frac{1}{2}\int_{-a}^a dy\sqrt{2V(y)} = \frac{S_o}{2}. \tag{2.9.19}$$

Close to the bottom of the well we can replace $V(x)$ by $\frac{(a-x)^2}{2}$. Hence,

$$\int_a^x dy\sqrt{2V(y)} = \int_a^x dy(a-y) = -\frac{(a-x)^2}{2}. \tag{2.9.20}$$

Thus, we have

$$\int_0^x dy\sqrt{2V(y)} = \frac{S_o}{2} - \frac{1}{2}(a-x)^2. \tag{2.9.21}$$

Recall that $\frac{dx}{dt} = \sqrt{2V(x)}$. Thus

$$\int_0^x dy\frac{E}{\sqrt{2V(y)}} = E\int_0^x dy\frac{1}{\frac{dy}{dt}}, \tag{2.9.22}$$

$$= E\int_0^t dt = Et. \tag{2.9.23}$$

Note that in the last line we choose $t = 0$ when $x = 0$, for simplicity. We are always free to choose the time $t = 0$ in any convenient way. The energy E is a small perturbation of the ground state energy of the oscillator

$$E = \frac{\hbar}{2} + \epsilon. \tag{2.9.24}$$

Recall that for large t, $\tilde{x}$ is close to a and we have

$$Ae^{-t} = \frac{1}{\sqrt{S_o}}\sqrt{2V(x)} = \frac{a-x}{\sqrt{S_o}}. \tag{2.9.25}$$

Thus, we have

$$t = \log\left(\frac{A\sqrt{S_o}}{a-x}\right). \tag{2.9.26}$$

Consequently,

$$Et = \left(\frac{\hbar}{2} + \epsilon\right)\log\left(\frac{A\sqrt{S_o}}{a-x}\right). \tag{2.9.27}$$

Finally, combining the above results, we have a complete analytic expression for the integral

$$\int_0^x dyk(y) = \frac{S_o}{2} - \frac{1}{2}(a-x)^2 - \left(\frac{\hbar}{2} + \epsilon\right)\log\left(\frac{A\sqrt{S_o}}{a-x}\right), \tag{2.9.28}$$

in the region close to $x = a$.

The WKB wavefunction of definite parity is

$$\Psi_\pm(x) = \frac{1}{\sqrt{k(x)}}\left[e^{\frac{1}{\hbar}\int_0^x dyk(y)} \pm e^{-\frac{1}{\hbar}\int_0^x dyk(y)}\right] \tag{2.9.29}$$

which, after using (2.9.28) becomes

$$\Psi_\pm(x) = \frac{1}{\sqrt{k(x)}}\left[e^{\frac{S_o}{2\hbar} - \frac{(a-x)^2}{2\hbar}}\sqrt{\frac{a-x}{A\sqrt{S_o}}} \pm e^{-\frac{S_o}{2\hbar} + \frac{(a-x)^2}{2\hbar}}\sqrt{\frac{A\sqrt{S_o}}{a-x}}\right], \tag{2.9.30}$$

Further, if we make the approximation

$$\sqrt{k(x)} = \sqrt{2V(x)} - \frac{E}{\sqrt{2V(x)}} \approx \sqrt{2V(x)}, \tag{2.9.31}$$

$$= \sqrt{a-x}, \tag{2.9.32}$$

the final expression for the wavefunction is

$$\Psi_{\pm}(x) = \frac{1}{\sqrt{a-x}} \left[e^{\frac{S_o}{2\hbar} - \frac{(a-x)^2}{2\hbar}} \sqrt{\frac{a-x}{A\sqrt{S_o}}} \pm e^{-\frac{S_o}{2\hbar} + \frac{(a-x)^2}{2\hbar}} \sqrt{\frac{A\sqrt{S_o}}{a-x}} \right]. \tag{2.9.33}$$

This wavefunction is clearly not normalizable. This is an artifact of our approximation and for very large x we know this wavefunction will receive further corrections that ultimately do give a normalizable wavefunction. Now consider the regions near the bottom of the potential. These regions are treated as a simple harmonic oscillator. Thus, the ground state wavefunction is

$$\psi_1(x) = e^{-\frac{(a-x)^2}{2\hbar}}. \tag{2.9.34}$$

Since the Schrödinger equation is a second order differential equation we know there is a second solution ϕ_1. To obtain ϕ_1 we will again use the Wronskian equation

$$\phi_1(x)\frac{d\psi_(x)}{dx} - \psi_1(x)\frac{d\phi_(x)}{dx} = C, \tag{2.9.35}$$

where C is a constant. In fact, choosing $\phi_1(x) = \frac{e^{\frac{(a-x)^2}{2\hbar}}}{a-x}$, and assuming

$$\frac{1}{(a-x)^2} \ll \frac{1}{\hbar} \tag{2.9.36}$$

we find $C = \frac{2}{\hbar}$. The proof is by direct calculation.

Proof.

$$\phi_1(x)\frac{d\psi_1(x)}{dx} - \psi_1(x)\frac{d\phi_(x)}{dx} = \frac{e^{\frac{(a-x)^2}{2\hbar}}}{a-x}\left(\frac{(a-x)}{2\hbar}e^{-\frac{(a-x)^2}{\hbar}}\right) + e^{-\frac{(a-x)^2}{2\hbar}}\left(\frac{e^{\frac{(a-x)^2}{2\hbar}}}{\hbar} - \frac{e^{\frac{(a-x)^2}{2\hbar}}}{(a-x)^2}\right),$$

$$= \frac{2}{\hbar} - \frac{1}{(a-x)^2}.$$

Clearly the last line gives $C = \frac{2}{\hbar}$, after use of the assumption in (2.9.36). □

Our wavefunction should be a small perturbation of the harmonic oscillator wavefunction. The time independent Schrödinger equation is

$$\left(\frac{-\hbar^2}{2}\frac{d^2}{dx} + \frac{(a-x)^2}{2}\right)\Psi = \left(\frac{\hbar}{2} + \epsilon\right)\Psi. \tag{2.9.37}$$

Introduce $\Psi_o(x)$ which satisfies

$$\left(-\frac{\hbar^2}{2}\frac{d^2}{dx} + \frac{(a-x)^2}{2} - \frac{\hbar}{2}\right)\Psi_o = 0. \tag{2.9.38}$$

In terms of the Green's function $G(x,x')$, we have

$$\Psi(x) = \Psi_o(x) + \epsilon \int dx' G(x,x')\Psi(x'). \tag{2.9.39}$$

The Green's function $G(x,x')$ obeys

$$\left(-\frac{\hbar^2}{2}\frac{d^2}{dx^2} + \frac{(a-x)^2}{2} - \frac{\hbar}{2}\right) G(x,x') = \delta(x-x'). \tag{2.9.40}$$

To see that (2.9.39) is a solution, note that

$$\left(-\frac{\hbar}{2}\frac{d^2}{dx^2} + \frac{(a-x)^2}{2} - \frac{\hbar^2}{2}\right)\Psi(x) = \left(H - \frac{\hbar}{2}\right)\Psi(x) = \epsilon\Psi(x), \tag{2.9.41}$$

$$\Leftrightarrow \left(H - \frac{\hbar}{2}\right)\left(\Psi_o(x) + \epsilon\int dx' G(x,x')\Psi(x')\right) = \epsilon\Psi(x), \tag{2.9.42}$$

$$\Rightarrow \left(H - \frac{\hbar}{2}\right) G(x,x') = \delta(x-x'). \tag{2.9.43}$$

It is not difficult to see that

$$G(x,x') = N[\theta(x-x')\phi_1(x)\psi_1(x') + \theta(x'-x)\phi_1(x')\psi_1(x)]. \tag{2.9.44}$$

This is indeed how we proceeded when we were working out the analytic expression for K in section 2.6. Integration of (2.9.43) yields

$$-\frac{\hbar^2}{2}\frac{dG}{dx}\Big|_{x=x'-\frac{\epsilon}{2}}^{x=x'+\frac{\epsilon}{2}} = \frac{\hbar^2}{2}N(\phi_1\frac{d\psi_1}{dx} - \psi_1\frac{d\phi_1}{dx}) = \hbar N = 1. \tag{2.9.45}$$

Hence, we have $N = \frac{1}{\hbar}$. Therefore in the region close to a such that $x' < x$, we have the following expression for Ψ.

$$\Psi(x) = e^{-\frac{(a-x)^2}{2\hbar}} + \frac{\epsilon}{\hbar}\int dx'\phi_1(x)\psi_1^2(x'). \tag{2.9.46}$$

Using the integral

$$\int dx'\psi_1^2(x') = \int dx' e^{-\frac{(a-x)^2}{\hbar}} = \sqrt{\pi\hbar}, \tag{2.9.47}$$

we obtain the final expression for Ψ at the bottom of the potential

$$\Psi(x) = e^{-\frac{(a-x)^2}{2\hbar}} + \epsilon\frac{e^{-\frac{(a-x)^2}{2\hbar}}}{a-x}\sqrt{\frac{\pi}{\hbar}}. \tag{2.9.48}$$

Matching the two wavefunctions (2.9.33) and (2.9.48), we can identify

$$Ae^{-S_o/\hbar}\sqrt{S_o} = \epsilon\sqrt{\frac{\pi}{\hbar}} \Rightarrow \epsilon = e^{-\frac{S_o}{\hbar}}\sqrt{\frac{\hbar S_o}{\pi}}A. \tag{2.9.49}$$

This is our final result for the value of the small perturbation of energy. This reproduces the level splitting computation using instantons perfectly.

2.10 Bounces

Bounces are another type of solution that extremize the Euclidean action. We could say that bounces are cousins of instantons. One example where bounces play a role is for the potential plotted in Figure 2.18 below

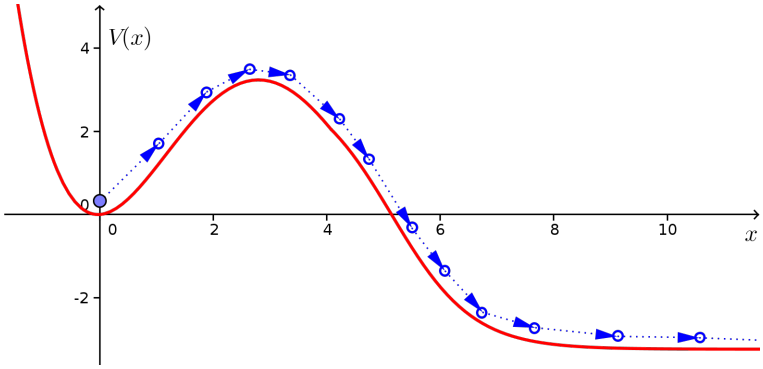


Figure 2.18: A particle tunnels from $x = 0$ through the potential barrier.

The state in which the particle sits at the bottom of the potential is unstable due to tunneling. Using the bounce we will be able to give a semiclassical description of this instability.

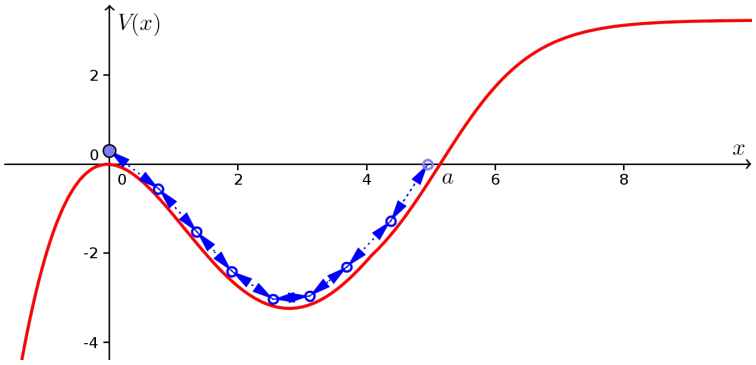


Figure 2.19: The Euclidean action has potential $-V(x)$. One possible Euclidean motion corresponds to a particle "bouncing" as shown.

From the Figure 2.21 we see that there is a motion in which the particle starts from $x = 0$, it rolls in the positive x direction, up the hill and then reverses direction to again come to rest at $x = 0$. This gives the bounce solution. The bounce is plotted in Figure 2.20 below.

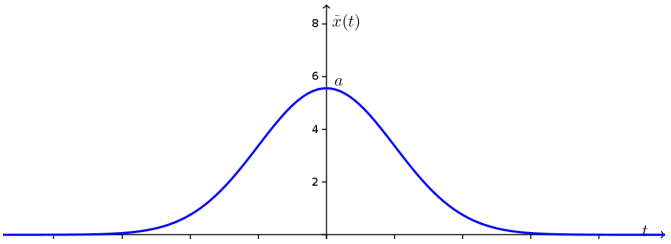


Figure 2.20: Spacetime diagram of the bounce.

The Euclidean action for a single bounce is

$$S_o = \int dt \left(\frac{1}{2} \left(\frac{dx}{dt} \right)^2 + V \right) = \int \sqrt{2V(x)}. \tag{2.10.1}$$

Using the techniques we developed in previous sections, we have

$$\langle 0|e^{-HT/\hbar}|0\rangle = \sum_i \frac{N e^{S[\tilde{x}_i/\hbar]}}{\sqrt{\det(-\partial_t^2 + V'')}} \tag{2.10.2}$$

In a dilute bounce gas approximation we have

$$\begin{aligned} \langle 0|e^{-HT/\hbar}|0\rangle &= \sum_n e^{-nS_o/\hbar} \sqrt{\frac{2\omega}{\pi\hbar}} e^{-\omega T/2} K^n \frac{T^n}{n!}, \\ &= \sqrt{\frac{2\omega}{\pi\hbar}} e^{-\omega T/2} e^{KTe^{-S_o/\hbar}}. \end{aligned}$$

Hence the expression for the energy is

$$E = \frac{\hbar\omega}{2} - \hbar K e^{-S_o/\hbar}. \tag{2.10.3}$$

There are a few comments that we can make about this formula. First, note that the bounce has corrected the energy E of the state. The correction is exponentially small in $\hbar$, which is much smaller than perturbative corrections would be. Why should we keep this correction when we have dropped perturbative corrections? Next, consider the value of K. Recall that K is given by

$$K = \sqrt{\prod_i \lambda_i} = \sqrt{\det(-\partial_t^2 + V'')}. \tag{2.10.4}$$

We know that we have a zero mode eigenstate $x_1 \propto \frac{d\tilde{x}}{dt}$ due to time translation invariance. Given the profile of $\tilde{x}(t)$ in Figure 2.20, the profile of the zero mode eigenstate is given in the following figure.

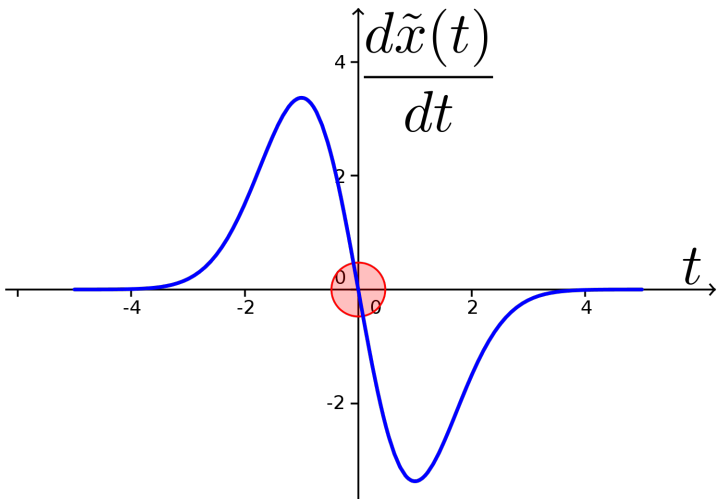


Figure 2.21: Profile of the zero mode associated to the time translation invariance. The node in the Euclidean wavefunction at $t = 0$ is shown.

This eigenfunction has a single node. Thus, there will be another eigenmode whose eigenfunction has no nodes. This eigenmode will have an even lower value of λ. Thus for the lowest mode we have $\lambda_0 < 0$. Since λ_0 is the only negative eigenvalue $\prod_i \lambda_i$ is negative. As a consequence K is pure imaginary. In total the small term which perturbs the energy in (2.10.3) is a pure imaginary. This observation is good news! Indeed, it explains why we keep the minute correction due to the bounce: this small correction is the leading contribution to the imaginary part of the energy of the state. Moreover, since we are getting a complex energy we know that we are not studying an energy eigenstate of the Hamiltonian. Indeed we know that our state is unstable so it is not an eigenstate of the full Hamiltonian. Finally, the imaginary contribution to the energy sets the lifetime of our particle. Indeed, note that

$$e^{-iEt/\hbar} = e^{-i\omega t/2 - |Ke^{-S_o/\hbar}|t}.$$

This wavefunction tell us that at large t the probability of finding the particle goes to zero, which is what we expect for an unstable state.

This concludes our discussion of instantons in quantum mechanics and provides the background for our study of instantons in Yang-Mills theories. Before we take this next step, we will review the quantization of Yang-Mills theories, in the next chapter.

3. Quantization of Gauge Theories

3.1 Why do we need "gauge symmetry"?

The success of Quantum Mechanics (QM) and Special Relativity (SR) are not only based on the beauty of the theories but also on their perfect agreement with experiment. At the time of writing, we know that theories of physics are necessarily quantum mechanical at very small length scales. Similarly, when the speed of particles become comparable to the speed of light, we need to use SR. Hence, both QM and SR have their own different domains of applicability in nature. These two domains can overlap, in which case our description must be both quantum mechanical and special relativistic. Examples of phenomena which need both QM and SR are the scattering experiments carried out at particle accelerators. As we will see, in constructing descriptions for particles with spin, there is a tension between QM and SR. Ultimately this tension is resolved by gauge symmetry.

3.2 The trouble with Canonical Quantization

In this section we will demonstrate the tension present between QM and SR by quantizing a free vector field. As a starting point, consider the following Lagrangian density of a free scalar

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2. \tag{3.2.1}$$

The scalar field can be thought of as a real valued function, $\phi(\vec{x},t)$ attached to each point of the 4-dimensional Minkowski spacetime. The mode expansion for ϕ can be written as

$$\phi(\vec{x},t) = \int \frac{d^3k}{(2\pi)^3}\frac{1}{2\omega_{\vec{k}}}\left(a_{\vec{k}}e^{-ik\cdot x} + a^\star_{\vec{k}}e^{ik\cdot x}\right), \tag{3.2.2}$$

where

$$k\cdot x = \left(\sqrt{\vec{k}^2+m^2}\right)t - \vec{k}\cdot\vec{x} = \omega_{\vec{k}}t - \vec{k}\cdot\vec{x}.$$

Note that up to this point our treatment is classical. In canonical quantization one needs to compute the canonical momentum

$$\Pi(\vec{x},t) = \frac{\partial\mathcal{L}}{\partial\dot{\phi}} = \frac{\partial\phi(\vec{x},t)}{\partial t}, \tag{3.2.3}$$

$$= \int \frac{d^3k}{(2\pi)^3}\frac{1}{2\omega_{\vec{k}}}(-i\omega_{\vec{k}})\left(a_{\vec{k}}e^{-ik\cdot x} - a^\star_{\vec{k}}e^{ik\cdot x}\right) \tag{3.2.4}$$

and impose the commutation relation

$$[\phi(\vec{y},t),\Pi(\vec{x},t)] = i\delta(\vec{x}-\vec{y}). \tag{3.2.5}$$

Of course Π and ϕ become operators in the quantization process. Hence, $a^\star_{\vec{k}},\ a_{\vec{k}}$ become respectively $a^\dagger(k),\ a(k)$ and obey the commutation relation

$$[a_{\vec{k}}, a^\dagger_{\vec{p}}] = (2\pi)^3 2\omega_{\vec{k}}\delta(\vec{k}-\vec{p}). \tag{3.2.6}$$

Hence the energy of a single mode is proportional to $|a_{\vec{k}}|^2$ or more precisely, to the number operator $N \propto a^\dagger_{\vec{k}} a_{\vec{k}}$.

In the case of a free vector field, by analogy to the scalar field, we should have

$$A_\mu(\vec{x},t) = \int \frac{d^3k}{(2\pi)^3}\frac{1}{2\omega_{\vec{k}}}[a_\mu(\vec{k})e^{-ik\cdot x} + a^\star_\mu(\vec{k})e^{ik\cdot x}], \tag{3.2.7}$$

where each component A_μ of the vector field is treated as a single scalar field. After the process of quantization we expect

$$[a_\mu(\vec{k}), a^\dagger_\nu(\vec{p})] = (2\pi)^3 2\omega_{\vec{k}}\delta(\vec{k}-\vec{p})\delta_{\mu\nu}. \tag{3.2.8}$$

Consider now a quantum state such as

$$|\Psi\rangle = \int \frac{d^3k}{(2\pi)^3} f^\mu(\vec{k}) a^\dagger_\mu(\vec{k})|0\rangle.$$

According to quantum mechanics, the norm of the state $|\Psi\rangle$ represents a probability. Thus, we should have

$$\big||\Psi\rangle\big|^2 = \langle 0| \int \frac{d^3k d^3k'}{(2\pi)^6} f^\mu(\vec{k})^\star f^\nu(\vec{k}') a_\mu(\vec{k}) a^\dagger_\nu(\vec{k}')|0\rangle, \tag{3.2.9}$$

$$= \int \frac{d^3k d^3k'}{(2\pi)^6} f^\mu(\vec{k})^\star f^\nu(\vec{k}')\langle 0|[a_\mu(\vec{k}), a^\dagger_\nu(\vec{k}')]|0\rangle, \tag{3.2.10}$$

$$= \int \frac{d^3k}{(2\pi)^3} 2\omega_{\vec{k}}\delta_{\mu\nu} f^{\mu\star}(\vec{k}) f^\nu(\vec{k}) \geq 0. \tag{3.2.11}$$

Clearly the positivity of norms of states required by QM is obeyed by the relation (3.2.11). However, SR is spoiled, because $\delta_{\mu\nu}$ is not a Lorentz invariant. It is clear that, replacing $\delta_{\mu\nu}$ with $-\eta_{\mu\nu}$ gives an equation that is consistent with SR. However doing so leads to a conflict with QM. In fact, since $\eta_{\mu\nu}$, has signature $(+,-,-,-)$, the norm of $|\Psi\rangle$ is no longer a probability because

$$\int \frac{d^3k}{(2\pi)^3} 2\omega_{\vec{k}}\eta_{\mu\nu} f^{\mu\star}(\vec{k}) f^\nu(\vec{k}) \geq 0 \tag{3.2.12}$$

is no longer guaranteed. Indeed, we have

$$-\int \frac{d^3k}{(2\pi)^3} 2\omega_{\vec{k}}\eta_{00}|f^0(\vec{k})|^2 < 0. \tag{3.2.13}$$

Resolution.

We have exhibited the tension between QM and SR in the case of a free vector field. We now want to argue that gauge invariance resolves this tension. Recall that the Lagrangian density associated to the free vector field is

$$\mathcal{L} = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}, \text{ where } F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{3.2.14}$$

The field strength $F_{\mu\nu}$ is invariant under the following transformation:

$$A_\mu \to A'_\mu = A_\mu + \partial_\mu \chi. \tag{3.2.15}$$

This transformation is known as an *Abelian gauge* transformation (it is an example of local symmetry) of the field A_μ. The two fields A'_μ and A_μ describe the same physical state. Thus the number of degrees

of freedom is decreased. As a consequence the states with negative norm are removed, and the tension between QM and SR is resolved. In general in QFT, "gauge symmetry" is always introduced to remove negative norm states.

Return to the action

$$S = \int d^4x \mathcal{L} = \int d^4x \left(-\tfrac{1}{4} F_{\mu\nu} F^{\mu\nu}\right). \tag{3.2.16}$$

The conjugate momentum are

$$\Pi^\mu = \frac{\partial \mathcal{L}}{\partial \dot{A}_\mu} = -\tfrac{1}{2} F^{\alpha\beta} \frac{\partial F_{\alpha\beta}}{\partial(\partial_0 A_\mu)}, \tag{3.2.17}$$

$$= -\tfrac{1}{2} F^{\alpha\beta} \frac{\partial(\partial_\alpha A_\beta - \partial_\beta A_\alpha)}{\partial(\partial_0 A_\mu)}, \tag{3.2.18}$$

$$= -\tfrac{1}{2} F^{\alpha\beta} (\delta^0_\alpha \delta^\mu_\beta - \delta^0_\beta \delta^\mu_\alpha) \tag{3.2.19}$$

$$= -\tfrac{1}{2} F^{0\mu} + \tfrac{1}{2} F^{\mu 0}, \tag{3.2.20}$$

$$\Pi^\mu = F^{\mu 0}. \tag{3.2.21}$$

Clearly, from the definition of $F^{\mu\nu}$, we have $\Pi^0 = 0$.

3.3 Quantization of a constrained system

In the quantization of the free vector field, we require the following commutations relations

$$[\Pi^0(t,\vec{x}), A^0(t,\vec{y})] = i\delta(\vec{x}-\vec{y}); \tag{3.3.1}$$

$$[\Pi^i(t,\vec{x}), A^j(t,\vec{y})] = -i\delta^{ij}\delta(\vec{x}-\vec{y}). \tag{3.3.2}$$

Note that (3.3.1) is not consistent with $\Pi^0 = 0$. One way out of this contradiction is to change our method of quantization. Systems like the free vector field are known as "constrained systems". The constraint is the relation $\Pi^0 = 0$. In the next paragraphs we will see how quantization is carried out for a constrained system. Dirac was the first to propose a method of quantizing constrained systems, [16].

We will work in the Heisenberg picture. For a constrained system, which has Hamiltonian H, the equations of motion are

$$i\frac{dO}{dt} = [O, H], \tag{3.3.3}$$

$$\chi_i = 0. \tag{3.3.4}$$

(3.3.3) is just the usual Heisenberg equation for an operator O and (3.3.4) are the constraints for our system. Now, the rules according to Dirac are the following:

Step 1 Determine the conditions implied by the time evolution of the constraints, i.e.

$$0 = i\frac{d\chi_i}{dt} = [\chi_i, H]. \tag{3.3.5}$$

(3.3.5) will lead to one of the following cases:

i) The set of equations (3.3.5) is nonsense. Hence the problem is not sensible.

ii) The set of equations (3.3.5) is consistent and obeyed when $\chi_i = 0$.

iii) The set of equations (3.3.5) are consistent but even when (3.3.5) is obeyed they imply some new constraints $\phi_j|_{j=1,2,\ldots} = 0$. The new constraints are known as secondary constraints.

Step 2 Repeat step 1 but replacing (3.3.5) by the following equations

$$\begin{cases} [\chi_i, H] &= 0 \\ [\phi_j, H] &= 0. \end{cases} \tag{3.3.6}$$

Step 3 Keep repeating the step 1 until we have a complete set of constraints

$$\varphi_k = \{\chi_i, \phi_j\}\,. \tag{3.3.7}$$

- Step A. Divide the constraints into 2 classes, 1^{st} class and 2^{nd} class. If φ_k is a second class constraint, then $[\varphi_j, \varphi_k]|_{\forall\varphi_l=0} \neq 0$. If a constraint is not 2^{nd} class, it is 1^{st} class. Denoting second class constraints η_i and first class constraints τ_i we have

$$\begin{aligned} [\tau_i, \varphi_k]_{\varphi_k=0} = 0 \;&\leftarrow \text{ no obstacle to setting } \tau_i = 0, \\ [\eta_i, \eta_j]_{\varphi_k=0} \neq 0 \;&\rightarrow \text{ can't set } \eta_i = 0. \end{aligned}$$

- Step B. Set

$$\tau_i = 0,$$

and change the method of quantization by changing the commutation relations to

$$[A, B] \to [A, B]_D = [A, B] - \sum_{i,j} [A, \eta_i] M_{ij} [\eta_j, B], \text{ with } (M^{-1})_{ij} = [\eta_i, \eta_j].$$

Under the new commutation relation, (3.3.3) is not changed

$$i\frac{dO}{dt}\bigg|_{\forall\varphi_l=0} = [O, H]_D|_{\forall\varphi_l=0} = \left[[O, H] - \sum_{i,j} [O, \eta_i] M_{ij} [\eta_j, H]\right]\bigg|_{\forall\varphi_l=0} = [O, H]. \tag{3.3.8}$$

The third equality above follows because we know that $[\eta_j, H]|_{\forall\varphi_l=0} = 0$. Also, we have

$$\begin{aligned} [\eta_k, \eta_l]_D &= [\eta_k, \eta_l] - \sum_{i,j} [\eta_k, \eta_i] M_{ij} [\eta_j, \eta_l], \\ &= [\eta_k, \eta_l] - \delta_{kj} [\eta_j, \eta_l], \\ &= [\eta_k, \eta_l] - [\eta_k, \eta_l] = 0. \end{aligned}$$

At this stage Dirac's algorithm of quantizing a constrained system can be considered as a set of arbitrary but self consistent rules. However, if we use this algorithm in quantizing the free vector field described previously, we will see that all these steps are sensible and reproduce what we observe in experiment.

Now let us return to the problem of the quantization of the vector field A^μ, armed with Dirac's algorithm. Recall that the first step is to ensure the constraints $\chi_i = 0$ are preserved in time. This implies

$$[\chi_i, H] = 0.$$

Our first constraint is

$$\chi_1 = \Pi^0 = 0. \tag{3.3.9}$$

Next, we can choose

$$\chi_2 = A^0 = 0. \tag{3.3.10}$$

That is because Π^0 does not appear in H so that $[A^0, H] = 0$ implying A^0 is a constant. We will choose A_μ so that

$$\chi_3 = \vec{\nabla} \cdot \vec{A} = 0. \tag{3.3.11}$$

(3.3.11) can be satisfied by making use of the gauge transformation (3.2.15), and (3.3.10). This is achieved if we choose χ so that

$$\vec{\nabla} \cdot \vec{A}' = \vec{\nabla} \cdot \vec{A} - \vec{\nabla} \cdot \vec{\nabla}\chi = 0. \tag{3.3.12}$$

We must ensure that $\partial_0 \chi = 0$ so that we don't disturb $A_0 = 0$. Indeed, under the above gauge transformation

$$A_0' = A_0 + \partial_0 \chi. \tag{3.3.13}$$

Note that, since there is no matter, Maxwell's equations imply

$$\vec{\nabla} \cdot \vec{E} = -\vec{\nabla} \cdot \vec{\nabla} A^0 - \vec{\nabla} \cdot \dot{\vec{A}} = 0. \tag{3.3.14}$$

In our $A_0 = 0$ gauge we have

$$\vec{\nabla} \cdot \vec{E} = -\vec{\nabla} \cdot \frac{\partial \vec{A}}{\partial t} = -\frac{\partial}{\partial t} \vec{\nabla} \cdot \vec{A} = 0. \tag{3.3.15}$$

From (3.3.12) we see that χ will indeed be independent of time. Our choice $A_0 = 0$ and $\vec{\nabla} \cdot \vec{A} = 0$ is known as Coulomb gauge. Thus we have three constraints (3.3.9), (3.3.10) and (3.3.11). We need to compute the commutator of these constraints with the Hamiltonian H. Consider the Hamiltonian H,

$$H = \int d^3x \left(\Pi^\mu \dot{A}_\mu - \mathcal{L} \right), \tag{3.3.16}$$

$$= \int d^3x \left[F^{\mu 0}(\partial_0 A_\mu - \partial_\mu A_0) + \tfrac{1}{4} F_{\mu\nu} F^{\mu\nu} + F^{\mu 0} \partial_\mu A_0 \right], \tag{3.3.17}$$

$$= \int d^3x \left(\tfrac{1}{2} F^{i0} F^{i0} + \tfrac{1}{2} F^{ij} F^{ij} + F^{i0} \partial_i A^0 \right), \tag{3.3.18}$$

$$= \int d^3x \left(\tfrac{1}{2} \vec{\Pi} \cdot \vec{\Pi} + \tfrac{1}{2} F^{ij} F^{ij} + \vec{\Pi} \cdot \vec{\nabla} A^0 \right), \tag{3.3.19}$$

$$= \int d^3x \left(\tfrac{1}{2} \vec{E} \cdot \vec{E} + \tfrac{1}{2} \vec{B} \cdot \vec{B} + \vec{\Pi} \cdot \vec{\nabla} A^0 \right). \tag{3.3.20}$$

In the last line we have used the equation (3.2.21) which implies

$$\begin{aligned} F^{i0} = \vec{\Pi} &= \partial^i A^0 - \partial^0 A^i, \\ &= -\vec{\nabla} A^0 - \frac{d\vec{A}}{dt}, \\ &= \vec{E} \end{aligned}$$

and $F^{ij} = \epsilon^{ijk}B^k$. Now compute the time evolution of the χ_i, which is determined by the commutator of χ_i with H.

$$[\chi_1(\vec{x},t), H] = [\Pi^0(\vec{x},t), H], \tag{3.3.21}$$

$$[\Pi^0(\vec{x},t), H] = \left[\Pi^0(\vec{x},t), \int d^3y \vec{\Pi}(\vec{y},t)\cdot\vec{\nabla}_{\vec{y}}A^0(\vec{y},t)\right], \tag{3.3.22}$$

$$= \int d^3y \vec{\Pi}(\vec{y},t)\cdot\vec{\nabla}_{\vec{y}}\left[\Pi^0(\vec{x},t), A^0(\vec{y},t)\right], \tag{3.3.23}$$

$$= -i\int d^3y \vec{\Pi}(\vec{y},t)\cdot\vec{\nabla}_{\vec{y}}\delta(\vec{x}-\vec{y}) \tag{3.3.24}$$

$$= i\vec{\nabla}_{\vec{y}}\cdot\vec{\Pi}(\vec{y},t) = 0 \Leftrightarrow \vec{\nabla}\cdot\vec{E} = 0. \tag{3.3.25}$$

Above the passage from the third equation to the fourth equation used (3.3.1). Hence $\vec{\nabla}\cdot\vec{\Pi} = 0$ is a secondary constraint.

Next we have

$$[\chi_2, H] = [A^0(\vec{x},t), H], \tag{3.3.26}$$

$$= 0. \tag{3.3.27}$$

A^0 commutes with H because Π^0 does not appear in H, (3.3.20).

Finally we have

$$[\chi_3, H] = [\vec{\nabla}_{\vec{x}}\cdot\vec{A}(\vec{x},t), H], \tag{3.3.28}$$

$$= \int d^3y \tfrac{1}{2}\left[\partial_{x^i}A^i(\vec{x},t), \left(\Pi^j(\vec{y},t)\right)^2\right] + \int d^3y \left[\partial_{x^i}A^i(\vec{x},t), \Pi^j(\vec{y},t)\partial_{y^j}A^0(\vec{y},t)\right], \tag{3.3.29}$$

$$= -i\int d^3y \delta^{ij}\Pi^j(\vec{y},t)\partial_{x^i}\delta(\vec{x}-\vec{y}) - \int d^3y \delta^{ij}\partial_{x^i}\partial_{y^j}A^0(\vec{y},t)\delta(\vec{x}-\vec{y}), \tag{3.3.30}$$

$$= \vec{\nabla}\cdot\vec{\Pi} + \vec{\nabla}\cdot\vec{\nabla}A^0 = 0. \tag{3.3.31}$$

The passage from the second equation to the third equation used (3.3.2). The last equation is obeyed because of the primary constraints and the secondary constraint. In summary, the complete set of constraints is

$$\chi_1 = \Pi^0 = 0,$$
$$\chi_2 = A^0 = 0,$$
$$\chi_3 = \vec{\nabla}\cdot\vec{A} = 0,$$
$$\chi_4 = \vec{\nabla}\cdot\vec{\Pi} = 0.$$

We now need to classify the constraints into two classes. Recall that φ_i is a second class constraint if $[\varphi_i, \chi_j]_{\varphi_i=0} \neq 0$. Hence all our constraints $\chi_i|_{i=1,2,3,4}$ are 2^{nd} class. In fact we have

$$[\chi_1(\vec{x},t), \chi_2(\vec{y},t)] = -i\delta(\vec{x}-\vec{y}), \tag{3.3.32}$$

$$[\chi_3(\vec{x},t), \chi_4(\vec{y},t)] = -i\vec{\nabla}_{\vec{x}}\cdot\vec{\nabla}_{\vec{y}}\delta(\vec{x}-\vec{y}). \tag{3.3.33}$$

By following our algorithm carefully, we are now ready to compute the matrix M defined by

$$(M^{-1})_{ij} = [\chi_i, \chi_j]$$

We have

$$M^{-1} = \begin{pmatrix} 0 & [\chi_1, \chi_2] & 0 & 0 \\ [\chi_2, \chi_1] & 0 & 0 & 0 \\ 0 & 0 & 0 & [\chi_3, \chi_4] \\ 0 & 0 & [\chi_4, \chi_3] & 0 \end{pmatrix}, \tag{3.3.34}$$

$$= \begin{pmatrix} 0 & -i\delta(\vec{x}-\vec{y}) & 0 & 0 \\ i\delta(\vec{x}-\vec{y}) & 0 & 0 & 0 \\ 0 & 0 & 0 & -i\vec{\nabla}_{\vec{x}}\vec{\nabla}_{\vec{y}}\delta(\vec{x}-\vec{y}) \\ 0 & 0 & i\vec{\nabla}_{\vec{x}}\vec{\nabla}_{\vec{y}}\delta(\vec{x}-\vec{y}) & 0 \end{pmatrix}. \tag{3.3.35}$$

M is determined by

$$\int d^3y \sum_j [M(\vec{x}_1, \vec{y})]_{ij} [M^{-1}(\vec{y}, \vec{x}_2)]_{jk} = \delta_{ik}\delta(\vec{x}_1 - \vec{x}_2). \tag{3.3.36}$$

It is straightforward to see that

$$M = \begin{pmatrix} 0 & -i\delta(\vec{x}-\vec{y}) & 0 & 0 \\ i\delta(\vec{x}-\vec{y}) & 0 & 0 & 0 \\ 0 & 0 & 0 & if(\vec{x}, \vec{y}) \\ 0 & 0 & -if(\vec{x}, \vec{y}) & 0 \end{pmatrix}; \tag{3.3.37}$$

where the function f is defined as

$$f(\vec{x}, \vec{y}) = \int \frac{d^3k}{(2\pi)^3} \frac{e^{i\vec{k}\cdot(\vec{x}-\vec{y})}}{\vec{k}\cdot\vec{k}}. \tag{3.3.38}$$

To verify this expression for M, focus on a single matrix element of equation (3.3.36), since the rest are very similar.

Proof.

$$\begin{aligned} \int d^3y [M(\vec{x}_1, \vec{y})]_{3a} [M^{-1}(\vec{y}, \vec{x}_2)]_{a3} &= \int d^3y [M(\vec{x}_1, \vec{y})]_{34} [M^{-1}(\vec{y}, \vec{x}_2)]_{43}, \\ &= -\int d^3y f(\vec{x}_1, \vec{y}) \vec{\nabla}_{\vec{x}_2} \vec{\nabla}_{\vec{y}} \delta(\vec{x_2} - \vec{y}), \\ &= \vec{\nabla}_{\vec{x}_2} \vec{\nabla}_{\vec{x}_1} f(\vec{x}_1, \vec{x}_2), \\ &= \vec{\nabla}_{\vec{x}_2} \vec{\nabla}_{\vec{x}_1} f(\vec{x}_1, \vec{x}_2) = \vec{\nabla}_{\vec{x}_2} \vec{\nabla}_{\vec{x}_1} \int \frac{d^3k}{(2\pi)^3} \frac{e^{i\vec{k}\cdot(\vec{x}_1-\vec{x}_2)}}{\vec{k}\cdot\vec{k}}; \\ &= \delta(\vec{x_1} - \vec{x}_2). \end{aligned}$$

□

We can now set $A^0 = \Pi^0 = 0$. Finally, we will compute the new commutators

$$\left[\vec{\Pi}(\vec{x}, t), \vec{A}(\vec{y}, t)\right]_D = \left[\vec{\Pi}(\vec{x}, t), \vec{A}(\vec{y}, t)\right] - \iint dz_1 dz_2 \sum_{k,l} [\Pi^i(\vec{x}, t), \chi_k(\vec{z_1}, t)] M_{kl}(z_1, z_2) [\chi_l(\vec{z_2}, t), \Pi^i(\vec{y}, t)].$$

Due to the relations (3.3.32) and (3.3.33), it is not difficult to see that the non-zero component of the second term in Dirac's commutator has $k = 3$ and $l = 4$. Hence

$$\left[\Pi^i(\vec{x},t), A^j(\vec{y},t)\right]_D = -i\delta^{ij}\delta(\vec{x}-\vec{y}) - \iint d^3z_1 d^3z_2 [\Pi^i(\vec{x},t), \chi_3(\vec{z}_1,t)] M_{34} [\chi_4(\vec{z}_2,t), A^j(\vec{y},t)], \tag{3.3.39}$$

$$= -i\delta^{ij}\delta(\vec{x}-\vec{y}) - \iint d^3z_1 d^3z_2 (i)\partial_{z_1^i}\delta(\vec{x}-\vec{z}_1)(i) f(\vec{z}_1,\vec{z}_2)(i)\partial_{z_2^j}\delta(\vec{y}-\vec{z}_2), \tag{3.3.40}$$

$$= -i\delta^{ij}\delta(\vec{x}-\vec{y}) + i\partial_{x^i}\partial_{y^i} f(\vec{x},\vec{y}), \tag{3.3.41}$$

$$= -i\int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot(\vec{x}-\vec{y})}\left(\delta^{ij} - \frac{k^i k^j}{\vec{k}\cdot\vec{k}}\right). \tag{3.3.42}$$

Introduce

$$\delta^{ij}_{\text{tr}} = \left(\delta^{ij} - \frac{k^i k^j}{\vec{k}\cdot\vec{k}}\right).$$

Construct two vectors $\vec{n}_1 \neq 0$ and $\vec{n}_2 \neq 0$ such that

$$\vec{k}\cdot\vec{n}_1 = \vec{k}\cdot\vec{n}_2 = \vec{n}_1\cdot\vec{n}_2 = 0. \tag{3.3.43}$$

It is easy to see that

$$\delta^{ij}_{\text{tr}} k^i = \left(\delta^{ij} - \frac{k^i k^j}{\vec{k}\cdot\vec{k}}\right) k^i = k^i - k^i = 0; \tag{3.3.44}$$

$$\delta^{ij}_{\text{tr}} n^j_l = \left(\delta^{ij} - \frac{k^i k^j}{\vec{k}\cdot\vec{k}}\right) n^j_l \Big|_{l=1,2} = n^j_l. \tag{3.3.45}$$

These two equations tell us about the polarization of the free vector field. We have exactly two polarizations, both perpendicular to the momentum of the particle. This is indeed what we expect.

Recall the commutation relation we obtained above

$$[\Pi^i(\vec{x},t), A^j(\vec{y},t)]_D = -i\int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot(\vec{x}-\vec{y})}\delta^{ij}_{\text{tr}}. \tag{3.3.46}$$

This equation can be used to determine the commutation relation between the electric field and the magnetic field. Concretely, we have

$$[E^i(\vec{x},t), B^j(\vec{y},t)]_D = [E^i(\vec{x},t), \epsilon^{jkl}\nabla_k A^l(\vec{y},t)]_D, \tag{3.3.47}$$

$$= -i\epsilon^{jkl}\partial_{y^k}\int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot(\vec{x}-\vec{y})}\left(\delta^{il} - \frac{k^j k^l}{\vec{k}\cdot\vec{k}}\right), \tag{3.3.48}$$

$$= -i\epsilon^{jki}\partial_{y^k}\delta(\vec{x}-\vec{y}) - \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot(\vec{x}-\vec{y})}\left(\frac{k^j k^i k^l \epsilon^{ikl}}{\vec{k}\cdot\vec{k}}\right), \tag{3.3.49}$$

$$= -i\epsilon^{jki}\partial_{y^k}\delta(\vec{x}-\vec{y}). \tag{3.3.50}$$

The passage to the last equation has used the antisymmetric property of ϵ^{jki} and the symmetric property of the product $k^j k^i k^l$. Finally we have shown that E^i and B^i do not commute with each other. By analogy to the Heisenberg uncertainty relation, we have

$$|\Delta\vec{E}_{\vec{k}}||\Delta\vec{B}_{\vec{k}}| \geq |\frac{\vec{k}}{2}|. \tag{3.3.51}$$

This inequality is very important because it is the only explanation of why atoms in excited states decay. It implies that there are always small fluctuations in the electric and magnetic fields and these small fluctuations perturb excited states into making transition to the ground state. The fluctuations are often interpreted as virtual particles popping in and out of the quantum vacuum.

3.4 Second example of Dirac quantization

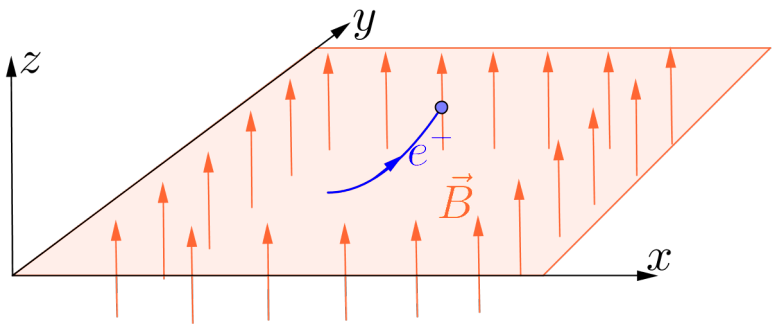


Figure 3.1: Electron immersed in an uniform magnetic field perpendicular to the $(x, y)-$plane.

In this section, as illustrated above, we will study the quantization of an electron moving in the $(x, y)-$plane and immersed in a uniform magnetic field perpendicular to the plane. Our goal is to obtain a deeper understanding of Dirac's constrained quantization. Clearly if the electron is not coupled to the vector field $\vec{B}$ the Hamiltonian of the system is just the Hamiltonian of a free particle

$$H_o = \frac{\vec{p}\cdot\vec{p}}{2m} = \frac{1}{2m}(p_x^2 + p_y^2).$$

The new Hamiltonian after coupling the electron and the magnetic vector potential $\vec{A}$ is given by

$$H = \frac{1}{2m}(\vec{p} - e\vec{A})\cdot(\vec{p} - e\vec{A}). \tag{3.4.1}$$

Note that this Hamiltonian is determined so that it is consistent with Newton's equation $\frac{d\vec{p}_{\text{total}}}{dt} = e\frac{d\vec{x}}{dt}\times\vec{B}$.

For a uniform field

$$\vec{B} = B\vec{e}_z = \vec{\nabla}\times\vec{A},$$

we take

$$\vec{A} = \begin{pmatrix} \frac{-B}{2}y \\ \frac{B}{2}x \\ 0 \end{pmatrix}. \tag{3.4.2}$$

Substituting the expression for $\vec{A}$ in (3.4.1) with its expression in (3.4.2) yields

$$H = \frac{1}{2m}\left[\left(p_x + \frac{\mathbf{e}By}{2}\right)^2 + \left(p_y - \frac{\mathbf{e}Bx}{2}\right)^2\right]. \tag{3.4.3}$$

This Hamiltonian looks like the Hamiltonian of a simple harmonic oscillator. Thus, by analogy with the harmonic oscillator, (3.4.3) can be rewritten in terms of the following operators.

$$a = \left(p_x + \frac{\mathbf{e}By}{2}\right) + i\left(p_y - \frac{\mathbf{e}Bx}{2}\right), \tag{3.4.4}$$

$$a^\dagger = \left(p_x + \frac{\mathbf{e}By}{2}\right) - i\left(p_y - \frac{\mathbf{e}Bx}{2}\right). \tag{3.4.5}$$

These particular choices imply

$$H = \frac{a^\dagger a + aa^\dagger}{4m}. \tag{3.4.6}$$

Direct calculation shows that

$$aa^\dagger = \left[\left(p_x + \frac{\mathbf{e}By}{2}\right) + i\left(p_y - \frac{\mathbf{e}Bx}{2}\right)\right]\left[\left(p_x + \frac{\mathbf{e}By}{2}\right) - i\left(p_y - \frac{\mathbf{e}Bx}{2}\right)\right], \tag{3.4.7}$$

$$= 2mH - i\frac{\mathbf{e}B}{2}\big([p_x, x] + [p_y, y]\big), \tag{3.4.8}$$

$$= 2mH + \mathbf{e}B\hbar; \tag{3.4.9}$$

$$a^\dagger a = \left[\left(p_x + \frac{\mathbf{e}By}{2}\right) - i\left(p_y - \frac{\mathbf{e}Bx}{2}\right)\right]\left[\left(p_x + \frac{\mathbf{e}By}{2}\right) + i\left(p_y - \frac{\mathbf{e}Bx}{2}\right)\right], \tag{3.4.10}$$

$$= 2mH + i\frac{\mathbf{e}B}{2}\big([p_x, x] + [p_y, y]\big), \tag{3.4.11}$$

$$= 2mH - \mathbf{e}B\hbar. \tag{3.4.12}$$

Above, we have used the usual commutation relations $[x, p_x] = [y, p_y] = i\hbar$. Thus, (3.4.6) can be obtained after summing (3.4.12) and (3.4.9). The advantage of computing the above expressions separately is that we can read the commutation relation between a and $a^\dagger$ directly from (3.4.12) and (3.4.9). We find

$$[a, a^\dagger] = 2\mathbf{e}B\hbar. \tag{3.4.13}$$

Thus, the spectrum of the system turns out to be the spectrum of the 1-dimensional harmonic oscillator. Indeed, defining the number operator $N = \frac{a^\dagger a}{2\mathbf{e}B\hbar}$ and using the commutation relation (3.4.13), we can rewrite the Hamiltonian as

$$H = \frac{1}{4m}\left(aa^\dagger + a^\dagger a\right) = \hbar\omega_c(N + \frac{1}{2}), \tag{3.4.14}$$

where $\omega_c = \frac{\mathbf{e}B}{m}$. Hence for any quantum state $|n\rangle$ for which $N|n\rangle = n|n\rangle$, we have

$$H|n\rangle = \hbar\omega_c(n + \frac{1}{2})|n\rangle. \tag{3.4.15}$$

Clearly the energy spectrum of the system is

$$E_n = \hbar\omega_c(n + \frac{1}{2}). \tag{3.4.16}$$

Our system is a two dimension system. However (3.4.16) seems to match the spectrum of a one dimensional system. Indeed, (3.4.16) is the energy spectrum for the 1-D harmonic oscillator. There is no real puzzle though because it turns out that each level in our system has an infinite degeneracy. For

a given quantum number "n", the energy $E_n = \hbar\omega_c(n+\frac{1}{2})$ is associated to the wavefunction $\psi_{n,l}$, with l a degeneracy label. In particular, the wavefunction of the ground state is given by

$$\psi_{0,l} = \sqrt{\left(\frac{\mathbf{e}B}{2\hbar}\right)^{l+1}}\sqrt{\frac{1}{l!\pi}}(x+iy)^l e^{-\frac{\mathbf{e}B}{4\hbar}(x^2+y^2)}. \tag{3.4.17}$$

The gap between two successive energy levels E_n is given by $\hbar\omega_c = \hbar\frac{\mathbf{e}B}{m}$. E_n are known as the Landau energy levels.

We now imagine that B is very big. We will only study the ground states. For large B the amount of energy require to go from the ground state to an excited state is so large that any small perturbations leave us in the ground state. Concretely our system is in a state $|\phi\rangle$, satisfying the following equation

$$a|\phi\rangle = 0. \tag{3.4.18}$$

Hence, we can take as our first constraint

$$\chi_1 = a = 0. \tag{3.4.19}$$

By daggering, the second constraint is easily obtained from this

$$\chi_2 = a^\dagger = 0. \tag{3.4.20}$$

We will now carry out Dirac's algorithm of quantization. Hence, we want to compute $[\chi_i, H]$. We have

$$[\chi_1, H] = [a, \frac{aa^\dagger + a^\dagger a}{4m}], \tag{3.4.21}$$

$$= \frac{1}{4m}\left(a[a,a^\dagger] + [a,a]a^\dagger + a^\dagger[a,a] + [a,a^\dagger]a\right), \tag{3.4.22}$$

$$= \hbar\omega_c a. \tag{3.4.23}$$

Similarly

$$[\chi_2, H] = [a^\dagger, H] = -\hbar\omega_c a^\dagger. \tag{3.4.24}$$

Equations (3.4.23) and (3.4.24) tell us that the only constraints are

$$\chi_1 = a = 0,$$
$$\chi_2 = a^\dagger = 0.$$

It is straightforward to see that these constraints are 2^{nd} class. Indeed, because of the relation (3.4.13), we have

$$[\chi_1, \chi_2] = 2\mathbf{e}B\hbar \neq 0. \tag{3.4.25}$$

The elements of the matrix $M^{-1}_{ij} = [\chi_i, \chi_j]$ can also be derived from (3.4.13). Thus, we have

$$M^{-1} = \begin{pmatrix} 0 & [\chi_1,\chi_2] \\ [\chi_2,\chi_1] & 0 \end{pmatrix} = \begin{pmatrix} 0 & 2\mathbf{e}B\hbar \\ -2\mathbf{e}B\hbar & 0 \end{pmatrix}. \tag{3.4.26}$$

The inverse of this matrix is not difficult to find. Indeed, we easily find

$$M = \begin{pmatrix} 0 & \frac{-1}{2\mathbf{e}B\hbar} \\ \frac{1}{2\mathbf{e}B\hbar} & 0 \end{pmatrix}. \tag{3.4.27}$$

Now we have all the ingredients that we need to compute the Dirac commutator. It will be convenient to consider the following operators

$$z = x + iy;$$
$$\bar{z} = x - iy.$$

Using the definition of the Dirac commutator, the commutation relation between these new operators is given by

$$[z,\bar{z}]_D = [z,\bar{z}] - \sum_{i,j}[z,\chi_i]M_{ij}[\chi_j,\bar{z}], \tag{3.4.28}$$
$$= 0 - [(x+iy),a]\left(\tfrac{-1}{2\mathbf{e}B\hbar}\right)[a^\dagger,(x-iy)] - [(x+iy),a^\dagger]\left(\tfrac{1}{2\mathbf{e}B\hbar}\right)[a,(x-iy)], \tag{3.4.29}$$
$$= -\tfrac{1}{2\mathbf{e}B\hbar}\left[x+iy,(p_x+\tfrac{\mathbf{e}By}{2})+i\left(p_y-\tfrac{\mathbf{e}Bx}{2}\right)\right]\left[(p_x+\tfrac{\mathbf{e}By}{2})-i\left(p_y-\tfrac{\mathbf{e}Bx}{2}\right),x-iy\right]$$
$$-\tfrac{1}{2\mathbf{e}B\hbar}\left[x+iy,(p_x+\tfrac{\mathbf{e}By}{2})-i\left(p_y-\tfrac{\mathbf{e}Bx}{2}\right)\right]\left[(p_x+\tfrac{\mathbf{e}By}{2})+i\left(p_y-\tfrac{\mathbf{e}Bx}{2}\right),x-iy\right], \tag{3.4.30}$$
$$= -\frac{1}{2\mathbf{e}B\hbar}\left([x,p_x]+[y,p_y]\right)\left([p_x,x]+[p_y,y]\right), \tag{3.4.31}$$
$$= -\frac{2\hbar}{\mathbf{e}B}. \tag{3.4.32}$$

We will reproduce this result in a second way that will clarify how Dirac's method is working. Towards this goal, study how z and $\bar{z}$ act on the ground state $\psi_{0,l}$. Recall that

$$\psi_{0,l} = \sqrt{\left(\frac{\mathbf{e}B}{2\hbar}\right)^{l+1}}\sqrt{\frac{1}{l!\pi}}(x+iy)^l e^{-\frac{\mathbf{e}B}{4\hbar}(x^2+y^2)}. \tag{3.4.33}$$

In terms of z and $\bar{z}$, we have

$$\psi_{0,l} = \sqrt{\left(\frac{\mathbf{e}B}{2\hbar}\right)^{l+1}}\sqrt{\frac{1}{l!\pi}}z^l e^{-\frac{\mathbf{e}B}{4\hbar}z\bar{z}}. \tag{3.4.34}$$

Hence, we read from this equation that the operator z acts on $\psi_{0,l}$ to give

$$z\psi_{0,l} = \sqrt{\frac{2\hbar(l+1)}{\mathbf{e}B}}\psi_{0,l+1}. \tag{3.4.35}$$

Equation (3.4.35) tells us that the operator z acting on $\psi_{0,l}$ increases l by one. This suggests that the operator $\bar{z}$ decreases l by one.

Define

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right), \tag{3.4.36}$$

it follows that

$$\frac{\partial z}{\partial z} = 1, \quad \frac{\partial \bar{z}}{\partial z} = 0.$$

As a consequence, we have

$$\left(\frac{\partial}{\partial z} + \frac{\mathbf{e}B\bar{z}}{4\hbar}\right)\psi_{0,l} = \sqrt{\frac{\mathbf{e}Bl}{2\hbar}}\psi_{0,l-1}. \tag{3.4.37}$$

Clearly the operator $\left(\frac{\partial}{\partial z}+\frac{\mathbf{e}B\bar{z}}{4\hbar}\right)$ is acting on $\psi_{0,l}$ to lower l by one. Next, consider the following integral

$$\int_{\mathbb{R}}dx\int_{\mathbb{R}}dy\psi^{\star}_{0,k}\bar{z}\psi_{0,l}=\left(\frac{\mathbf{e}B}{2\hbar}\right)^{\frac{k+l}{2}+1}\frac{1}{\sqrt{k!l!}\pi}\int_{\mathbb{R}}dx\int_{\mathbb{R}}dy\bar{z}^{k+1}z^{l}e^{-\frac{\mathbf{e}B}{2\hbar}z\bar{z}}, \tag{3.4.38}$$

$$=\left(\frac{\mathbf{e}B}{2\hbar}\right)^{\frac{k+l}{2}+1}\frac{1}{\sqrt{k!l!}\pi}\int_{0}^{2\pi}d\theta e^{i(l-k-1)\theta}\frac{2}{2}\int_{0}^{\infty}drr^{2\frac{(k+l+3)}{2}-1}e^{-\frac{\mathbf{e}B}{2\hbar}r^{2}}, \tag{3.4.39}$$

$$=\left(\frac{\mathbf{e}B}{2\hbar}\right)^{\frac{k+l}{2}+1}\frac{1}{\sqrt{k!l!}\pi}\int_{0}^{2\pi}d\theta e^{i(l-k-1)\theta}\left(\sqrt{\frac{2\hbar}{\mathbf{e}B}}\right)^{(l+k+3)}\frac{\Gamma(\frac{k+l+3}{2})}{2}, \tag{3.4.40}$$

$$=\sqrt{\frac{2\hbar}{\mathbf{e}B}}\frac{\Gamma(\frac{k+l+3}{2})}{2\pi\sqrt{k!l!}}\int_{0}^{2\pi}d\theta e^{i(l-k-1)\theta}. \tag{3.4.41}$$

Above we used the change of variable $z=x+iy=re^{i\theta}$ and the definition of the Euler Γ function. Equation (3.4.41) tells us that the only value for k which gives a non-zero result is $k=l+1$. Hence in the case where $k=l+1$, (3.4.41) becomes

$$\int_{\mathbb{R}}dx\int_{\mathbb{R}}dy\psi^{\star}_{0,l-1}\bar{z}\psi_{0,l}=\sqrt{\frac{2\hbar}{\mathbf{e}B}}\frac{l!2\pi}{2\pi(l-1)!\sqrt{l}}=\sqrt{\frac{2\hbar l}{\mathbf{e}B}}. \tag{3.4.42}$$

We can read from this last equation that

$$\bar{z}\psi_{0,l}=\sqrt{\frac{2\hbar l}{\mathbf{e}B}}\psi_{0,l-1}. \tag{3.4.43}$$

Hence comparing equations (3.4.37) and (3.4.43) we can relate $\bar{z}$ and $\frac{\partial}{\partial z}+\frac{\mathbf{e}B\bar{z}}{4\hbar}$ as

$$\sqrt{\frac{\mathbf{e}B}{2\hbar l}}\bar{z}=\sqrt{\frac{2\hbar}{\mathbf{e}Bl}}\left(\frac{\partial}{\partial z}+\frac{\mathbf{e}B\bar{z}}{4\hbar}\right); \tag{3.4.44}$$

$$\bar{z}=\frac{2\hbar}{\mathbf{e}B}\left(\frac{\partial}{\partial z}+\frac{\mathbf{e}B\bar{z}}{4\hbar}\right). \tag{3.4.45}$$

Regarding equation (3.4.45), $\bar{z}$ and $\frac{2\hbar}{\mathbf{e}B}\left(\frac{\partial}{\partial z}+\frac{\mathbf{e}B\bar{z}}{4\hbar}\right)$ have the same matrix elements between lowest Landau level states, but the second operator additionally also leaves the lowest Landau level invariant, whereas the first doesn't. From this equation we have the commutation relation

$$[z,\bar{z}]=-\frac{2\hbar}{\mathbf{e}B}. \tag{3.4.46}$$

This is precisely the answer we found from Dirac's constrained quantization method. This is the understanding we were after.

In terms of x and y we have

$$[x+iy,x-iy]=-\frac{2\hbar}{\mathbf{e}B}; \tag{3.4.47}$$

$$[x,y]=\frac{\hbar}{i\mathbf{e}B}. \tag{3.4.48}$$

The Heisenberg uncertainty relation that can be derived from this commutator implies that

$$\Delta x\Delta y\geq\frac{\hbar}{2\mathbf{e}B}. \tag{3.4.49}$$

The physics behind this relation is illustrated in the following figure and explained below.

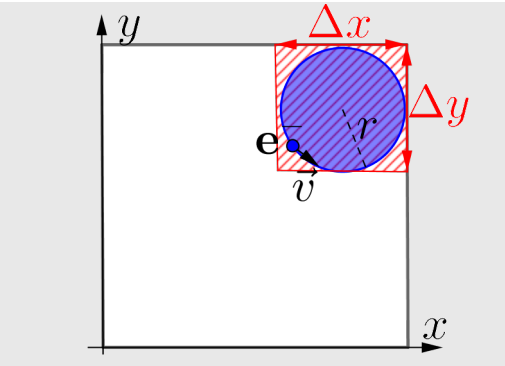


Figure 3.2: An electron in a background magnetic field occupies finite area.

Applying Newton's second law to an electron gives

$$\frac{mv^2}{r} = \mathbf{e}vB. \tag{3.4.50}$$

The left hand side of (3.4.50) is the mass times the radial acceleration, while the right hand side is the Lorentz force. It is straightforward to see that (3.4.50) can be rewritten such that

$$\frac{2\pi mvr}{\hbar} = \frac{2\mathbf{e}B\pi r^2}{\hbar};\ S = \pi r^2 \tag{3.4.51}$$

$$\Leftrightarrow$$

$$\frac{mvr}{h} = l = \frac{2\mathbf{e}BS}{\hbar}. \tag{3.4.52}$$

Above S is the area of the disc delimited by the trajectory of the electron and l is the total angular momentum quantum number of the electron. At the quantum level, the lowest possible value for a spinning electron is $l = 1$. Thus, we have

$$S = \frac{\hbar}{2\mathbf{e}B}. \tag{3.4.53}$$

Finally (3.4.49) can be rewritten as

$$\Delta x \Delta y \geq S = \pi r^2. \tag{3.4.54}$$

We can not measure x and y precisely because our circulating electron effectively takes up an area S.

In this section we have illustrated constrained quantization on a very simple system. This discussion has explained how constrained quantization works and has given some insight into what the Dirac brackets achieve. This sets the foundations for the system of interest to us, the free vector field, which is studied in the next section.

3.5 Coulomb gauge quantization

In this section we return to the free vector field in Coulomb gauge. The aim of this section is to perform in detail the calculation of the photon propagator in the Coulomb gauge. Indeed we will show that the

photon propagator in the Coulomb gauge is given by the following expression

$$iD_{\mu\nu}(x',x) \equiv \langle 0|T\left(A_\mu(x')A_\nu(x)\right)|0\rangle = \int \frac{d^4k}{(2\pi)^4}\frac{e^{-ik\cdot(x'-x)}}{k^2+i\epsilon}\sum_{\lambda=1}^{2}\xi_\mu(\vec{k},\lambda)\xi_\nu(\vec{k},\lambda). \tag{3.5.1}$$

Recall that the Coulomb gauge is obtained by imposing the following conditions on the vector potential A^μ,

$$\begin{aligned} A^0 &= 0; \\ \vec{\nabla}\cdot\vec{A} &= 0. \end{aligned}$$

In this gauge, Maxwell's equations are reduced to

$$\partial_\mu\partial^\mu\vec{A} = 0. \tag{3.5.2}$$

We showed in (3.3.45) that the solutions $\vec{A}$ have only two polarization vectors lying in the plane perpendicular to the momentum $\vec{k}$ of the field. A general initial condition for (3.5.2) is given by

$$\vec{A}(\vec{x}) \equiv \vec{A}(\vec{x},t=0) = \int \frac{d^3k}{(2\pi)^3}\frac{1}{2\omega}\sum_{\lambda=1}^{2}\vec{\xi}(\vec{k},\lambda)\left[\alpha(\vec{k},\lambda)e^{i\vec{k}\cdot\vec{x}} + \alpha^\dagger(\vec{k},\lambda)e^{-i\vec{k}\cdot\vec{x}}\right]. \tag{3.5.3}$$

It is not difficult to show that the time dependence of the vector field can be derived from the relation

$$\vec{A}(\vec{x},t) = e^{iHt}\vec{A}(\vec{x})e^{-iHt} \tag{3.5.4}$$

to yield

$$\vec{A}(\vec{x},t) \equiv \vec{A}(x) = \int \frac{d^3k}{(2\pi)^3}\frac{1}{2\omega}\sum_{\lambda=1}^{2}\vec{\xi}(\vec{k},\lambda)\left[\alpha(\vec{k},\lambda)e^{-ik\cdot x} + \alpha^\dagger(\vec{k},\lambda)e^{ik\cdot x}\right]. \tag{3.5.5}$$

Above, λ is not a parameter but an index running from 1 to 2, indicating the two polarizations. k is the four-momentum vector in the Minkowski spacetime. For a photon, we have the relation $k^2 = 0 \Rightarrow \omega = k^0 = |\vec{k}|$. A simple calculation shows that indeed

$$\vec{\nabla}\cdot\vec{A} = 0 \Rightarrow \vec{k}\cdot\vec{\xi}(\vec{k},\lambda) = 0.$$

Note that the choice for the two polarization vectors found in (3.3.43) can be written as

$$\vec{\xi}(\vec{k},\lambda)\cdot\vec{\xi}(\vec{k},\lambda') = \delta_{\lambda\lambda'}. \tag{3.5.6}$$

Since α and $\alpha^\dagger$ respectively play the role of a and $a^\dagger$, they satisfy the commutation relations

$$[\alpha(\vec{k},\lambda),\alpha^\dagger(\vec{k}',\lambda')] = (2\pi)^3 2\omega\delta(\vec{k}-\vec{k}')\delta_{\lambda\lambda'}; \tag{3.5.7}$$

$$[\alpha(\vec{k},\lambda),\alpha(\vec{k}',\lambda')] = [\alpha^\dagger(\vec{k},\lambda),\alpha^\dagger(\vec{k}',\lambda')] = 0. \tag{3.5.8}$$

These can be derived from (3.3.46). Now we want to express the Hamiltonian H in terms of the operators α and $\alpha^\dagger$. In the Coulomb gauge, the Hamiltonian is

$$H = \int d^3x\left(\frac{1}{2}(\partial^0 A^i)^2 + \frac{1}{2}(\partial^i A^j - \partial^j A^i)^2\right). \tag{3.5.9}$$

Substituting A^i with its expression in (3.5.5), we have

$$H = \sum_{\lambda=1}^{2} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2} \alpha^\dagger(\vec{k},\lambda)\alpha(\vec{k},\lambda). \tag{3.5.10}$$

We are now going to compute the photon propagator in Coulomb gauge. Towards this end, consider

$$\langle 0|A_\mu(x')A_\nu(x)|0\rangle = \langle 0| \sum_{\lambda,\lambda'} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2\omega'} \xi_\mu(\vec{k},\lambda') e^{-ik\cdot x'} \alpha(\vec{k},\lambda) \int \frac{d^3p}{(2\pi)^3} \frac{1}{2\omega} \xi_\nu(\vec{p},\lambda) e^{ip\cdot x} \alpha^\dagger(\vec{p},\lambda)|0\rangle, \tag{3.5.11}$$

$$= \sum_{\lambda,\lambda'} \int \frac{d^3k d^3p}{(2\pi)^6} \frac{1}{4\omega\omega'} \xi_\mu(\vec{k},\lambda')\xi_\nu(\vec{p},\lambda) e^{-ik\cdot x' + ip\cdot x} \langle 0|\alpha(\vec{k},\lambda')\alpha^\dagger(\vec{p},\lambda)|0\rangle. \tag{3.5.12}$$

Above we used the relation $\langle 0|A_\mu(x')A_\nu(x)|0\rangle \propto \langle 0|\alpha\alpha^\dagger|0\rangle$. Indeed this follows from

$$\alpha(\vec{p},\lambda)|0\rangle = 0,$$
$$\langle 0|\alpha^\dagger(\vec{k},\lambda') = 0.$$

Next, note that (3.5.12) can be rewritten using the commutator $[\alpha(\vec{k},\lambda'), \alpha^\dagger(\vec{p},\lambda)]$, again because of the above two equations. Thus,

$$\langle 0|A_\mu(x')A_\nu(x)|0\rangle = \sum_{\lambda,\lambda'} \int \frac{d^3k d^3p}{(2\pi)^6} \frac{1}{4\omega\omega'} \xi_\mu(\vec{k},\lambda')\xi_\nu(\vec{p},\lambda) e^{-ik\cdot x' + ip\cdot x} \langle 0| \left[\alpha(\vec{k},\lambda'), \alpha^\dagger(\vec{p},\lambda)\right] |0\rangle. \tag{3.5.13}$$

Using (3.5.7) we find

$$\langle 0|A_\mu(x')A_\nu(x)|0\rangle = \sum_{\lambda=1}^{2} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2\omega} \xi_\mu(\vec{k},\lambda)\xi_\nu(\vec{k},\lambda) e^{-ik\cdot(x'-x)}. \tag{3.5.14}$$

A very similar calculation shows that

$$\langle 0|A_\nu(x)A_\mu(x')|0\rangle = \sum_{\lambda=1}^{2} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2\omega} \xi_\mu(\vec{k},\lambda)\xi_\nu(\vec{k},\lambda) e^{ik\cdot(x'-x)}. \tag{3.5.15}$$

We define the time-ordered product of Heisenberg operators as

$$T\left(A_\mu(x')A_\nu(x)\right) = \theta(x^{0\prime} - x^0)A_\mu(x')A_\nu(x) + \theta(x^0 - x^{0\prime})A_\nu(x)A_\mu(x'). \tag{3.5.16}$$

The vacuum expectation value of $T\left(A_\mu(x')A_\nu(x)\right)$ with the help of (3.5.14) and (3.5.15), is

$$\langle 0|T\left(A_\mu(x')A_\nu(x)\right)|0\rangle = \sum_{\lambda=1}^{2} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2\omega} \xi_\mu(\vec{k},\lambda)\xi_\nu(\vec{k},\lambda) \Big(\theta(x^{0\prime} - x^0)e^{-ik\cdot(x'-x)} + \theta(x^0 - x^{0\prime})e^{ik\cdot(x'-x)}\Big)\Big|_{k^0=\omega} \tag{3.5.17}$$

Consider now the integral

$$\int \frac{d^4k}{(2\pi)^4} \frac{e^{-ik\cdot(x'-x)}}{k^2 + i\epsilon} = \int \frac{d^3k}{(2\pi)^3} \int \frac{dk^0}{2\pi} \frac{e^{-ik\cdot(x'-x)}}{(k^0 - \omega + i\frac{\epsilon}{2\omega})(k^0 + \omega - i\frac{\epsilon}{2\omega})}. \tag{3.5.18}$$

The right hand side of this equation used the identity $(k^0)^2 - \omega^2 + i\epsilon = (k^0 - \omega + i\frac{\epsilon}{2\omega})(k^0 + \omega - i\frac{\epsilon}{2\omega})$, accurate up to terms of $O(\epsilon^2)$. The integral over k^0 can be performed using the residue theorem. By noting that

$$k^0 = k^0_r + ik^0_{im}, \quad \text{we have}$$
$$\Rightarrow -ik^0(x^{0\prime} - x^0) = -ik^0_r(x^{0\prime} - x^0) + k^0_{im}(x^{0\prime} - x^0).$$

From this last equation we see that, if the exponential in the integrand of (3.5.18) is to vanish, we need

$$\begin{cases} k^0_{im} > 0 & \Rightarrow x^{0\prime} < x^0; \\ k^0_{im} < 0 & \Rightarrow x^{0\prime} > x^0 \end{cases}.$$

Thus, if $x^{0\prime} - x^0 > 0$ the integral over k^0 in (3.5.18) is performed along the following closed path.

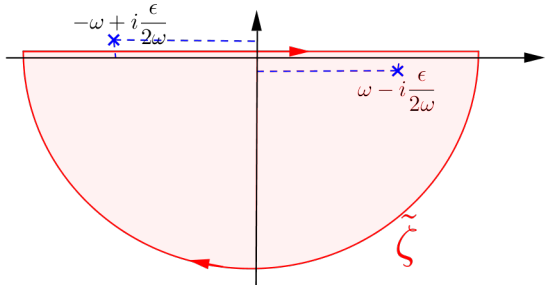


Figure 3.3: Picking up the pole under the real line

Thus, we have

$$\int \frac{d^4k}{(2\pi)^4} \frac{e^{-ik\cdot(x'-x)}}{k^2 + i\epsilon} = \int \frac{d^3k}{(2\pi)^4} \oint_{\tilde{\zeta}} \frac{dk^0}{2\pi} \frac{e^{-ik\cdot(x'-x)}}{(k^0 - \omega + i\frac{\epsilon}{2\omega})(k^0 + \omega - i\frac{\epsilon}{2\omega})}, \tag{3.5.19}$$
$$= \int \frac{d^3k}{(2\pi)^3}(-i)\frac{e^{-i\omega(x^{0\prime}-x^0)+i\vec{k}\cdot(\vec{x}'-\vec{x})}}{-\omega - \omega}. \tag{3.5.20}$$

Similarly, if $x^{0\prime} - x^0 < 0$ we must close in the upper half of the complex plane.

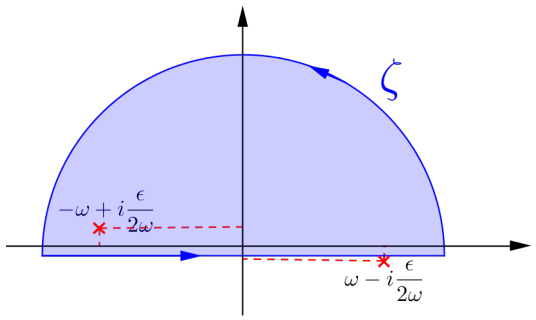


Figure 3.4: Picking up the pole above the real line

Now, we have

$$\int \frac{d^4k}{(2\pi)^4}\frac{e^{-ik\cdot(x'-x)}}{k^2+i\epsilon} = \int \frac{d^3k}{(2\pi)^4}\oint_\zeta \frac{dk^0}{2\pi}\frac{e^{-ik\cdot(x'-x)}}{(k^0-\omega+i\frac{\epsilon}{2\omega})(k^0+\omega-i\frac{\epsilon}{2\omega})}, \tag{3.5.21}$$

$$= \int \frac{d^3k}{(2\pi)^3}(i)\frac{e^{-i\omega(x^{0\prime}-x^0)+i\vec{k}\cdot(\vec{x}'-\vec{x})}}{-\omega-\omega}, \tag{3.5.22}$$

$$= \int \frac{d^3k}{(2\pi)^3}(-i)\frac{e^{i\omega(x^{0\prime}-x^0)-i\vec{k}\cdot(\vec{x}'-\vec{x})}}{2\omega}. \tag{3.5.23}$$

Combining (3.5.20) and (3.5.23), we have

$$\int \frac{d^4k}{(2\pi)^4}\frac{e^{-ik\cdot(x'-x)}}{k^2+i\epsilon} = \theta(x^{0\prime}-x^0)\int \frac{d^3k}{(2\pi)^3}\frac{(-i)}{2\omega}e^{-i\omega(x^{0\prime}-x^0)+i\vec{k}\cdot(\vec{x}'-\vec{x})} + \theta(x^0-x^{0\prime})\int \frac{d^3k}{(2\pi)^3}\frac{(-i)}{2\omega}e^{i\omega(x^{0\prime}-x^0)-i\vec{k}\cdot(\vec{x}'-\vec{x})} \tag{3.5.24}$$

Finally, the photon propagator is easily derived from this equation and thus

$$iD_{\mu\nu}(x',x) = \int \frac{d^4k}{(2\pi)^4}\frac{e^{-ik\cdot(x'-x)}}{k^2+i\epsilon}\sum_{\lambda=1}^{2}\xi_\mu(\vec{k},\lambda)\xi_\nu(\vec{k},\lambda). \tag{3.5.25}$$

3.6 Lorentz Gauge Quantization: Gupta-Bleuler Formalism

Dirac's constrained quantization is not manifestly Lorentz invariant. The Gupta-Bleuler formalism attempts to make Lorentz covariance manifest. As a starting point, consider a Lorentz invariant gauge fixing condition

$$\partial_\mu A^\mu = 0. \tag{3.6.1}$$

Under a gauge transformation

$$A'^\mu = A^\mu + \partial^\mu\chi, \tag{3.6.2}$$

we find that we remain in Lorentz gauge

$$\partial_\mu A'^\mu = \partial_\mu A^\mu = 0 + \partial_\mu\partial^\mu\chi = 0 \tag{3.6.3}$$

if

$$\partial_\mu\partial^\mu\chi = 0.$$

Thus the Lorentz gauge is only a partial gauge fixing. Recall that in the quantum mechanical theory of the vector field, the commutation relation between $A^\mu(\vec{x},t)$ and $\Pi^\nu(\vec{y},t)$ is

$$[A^\mu(\vec{x},t),\Pi^\nu(\vec{y},t)] = -i\eta^{\mu\nu}\delta(\vec{x}-\vec{y}). \tag{3.6.4}$$

Now, if we replace the original Lagrangian density $\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu}$ by

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} \to \mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} - \frac{\alpha}{2}(\partial_\mu A^\mu)^2 \tag{3.6.5}$$

in the Lorentz gauge the physics of the system does not change. It is straightforward to see that the equations of motion become

$$\partial^\mu F_{\mu\nu} + \partial^\mu \partial_\nu A_\mu = \partial^\mu \partial_\mu A_\nu - (1-\alpha)\partial_\nu \partial^\mu A_\mu = 0. \tag{3.6.6}$$

We still have to impose the gauge condition, (3.6.1). Consider

$$\partial_\mu A^\mu = 0, \tag{3.6.7}$$

which imposes the gauge condition as an operator equation. If we act with the operator $\frac{\partial}{\partial x^\mu}$ on the commutation relation (3.6.4) the resulting equation contradicts (3.6.7). In an attempt to avoid this tension, we try to impose the gauge fixing as the weaker state vector equation

$$\partial_\mu A^\mu |\psi_{\text{phys}}\rangle = 0. \tag{3.6.8}$$

Acting with the operator $\frac{\partial}{\partial x^\mu}$ on the commutation relation (3.6.4) and taking the expectation value in state $|\psi_{\text{phys}}\rangle$, we again find an inconsistency. Finally, imposing the gauge through the even weaker condition

$$\langle\psi_{\text{phys}}|\partial_\mu A^\mu|\psi_{\text{phys}}\rangle = 0, \tag{3.6.9}$$

there is no obvious contradiction with (3.6.4). For a specific value $\alpha = 1$, which is known as the Feynman gauge, the simplified version of the Lagrangian density is

$$\mathcal{L} = -\frac{1}{2}\partial_\nu A_\mu \partial^\nu A^\mu. \tag{3.6.10}$$

Now the conjugate momentum Π_μ are easily obtained

$$\Pi^\mu = \frac{\partial \mathcal{L}}{\partial(\partial_0 A_\mu)} = -\partial^0 A^\mu. \tag{3.6.11}$$

The equations of motion (3.6.6), in the Feynman gauge are

$$\partial^\mu \partial_\mu A_\nu = 0. \tag{3.6.12}$$

The Hamiltonian density of the vector field is computed by analogy with the calculations that we did to get (3.3.20). We find

$$\begin{aligned}\mathcal{H} &= \Pi^\mu \partial_0 A_\mu - \mathcal{L}, \\ &= -\frac{1}{2}(\Pi_0)^2 - \frac{1}{2}(\vec{\nabla}A_0)^2 + \frac{1}{2}\vec{\Pi}\cdot\vec{\Pi} + \frac{1}{2}\vec{\nabla}A^i \cdot \vec{\nabla}A^i.\end{aligned}$$

Note that this Hamiltonian density is indeed the Hamiltonian density of a collection of harmonic oscillators except that we have four oscillators and one of them has the wrong sign.

We now need to enforce (3.6.9). Recall that the commutation relation and the equations of motion are respectively:

$$\begin{gathered}[A^\mu(\vec{x},t), \Pi^\nu(\vec{y},t)] = -i\eta^{\mu\nu}\delta(\vec{x}-\vec{y}), \\ \partial^\mu\partial_\mu A_\nu = 0.\end{gathered}$$

Hence, the solution $A_\mu(x)$ can be written in terms of Fourier modes as

$$A_\mu(x) = \int \frac{d^3k}{(2\pi)^3}\frac{1}{2\omega}\sum_{\lambda=1}^{3}\left(\alpha(\vec{k},\lambda)\xi_\mu(\vec{k},\lambda)e^{-ik\cdot x} + \alpha^\dagger(\vec{k},\lambda)\xi^\star_\mu(\vec{k},\lambda)e^{ik\cdot x}\right), \tag{3.6.13}$$

$$= A^+_\mu(\alpha,x,\lambda) + A^-_\mu(\alpha^\dagger,x,\lambda). \tag{3.6.14}$$

We have introduced $A_\mu^\pm$ because they will simplify our calculations.

From the commutation relations that A_μ and Π_ν satisfy, we can argue that the following equations must be satisfied

$$\xi(\vec{k},\lambda)\cdot\xi^\star(\vec{k},\lambda') = \eta^{\mu\nu}\xi_\mu(\vec{k},\lambda)\xi_\nu(\vec{k},\lambda'), \tag{3.6.15}$$

$$= \eta_{\lambda\lambda'}; \tag{3.6.16}$$

$$\Big[\alpha(\vec{k},\lambda),\alpha^\dagger(\vec{p},\lambda')\Big] = -2\omega(2\pi)^3\delta(\vec{k}-\vec{p})\eta_{\lambda\lambda'}. \tag{3.6.17}$$

It is not difficult to see that the vectors $\xi(\vec{k},\lambda)$ can be chosen to be the canonical basis vectors of $\mathbb{R}^4$. Thus we have

$$\xi(\vec{k},\lambda=0)=\begin{pmatrix}1\\0\\0\\0\end{pmatrix},\quad \xi(\vec{k},\lambda=1)=\begin{pmatrix}0\\1\\0\\0\end{pmatrix},\quad \xi(\vec{k},\lambda=2)=\begin{pmatrix}0\\0\\1\\0\end{pmatrix},\quad \xi(\vec{k},\lambda=3)=\begin{pmatrix}0\\0\\0\\1\end{pmatrix}.$$

Now, if we use the relation (3.6.14) in the constraint $\langle\psi_{\text{phys}}|\partial_\mu A^\mu|\psi_{\text{phys}}\rangle = 0$, we have

$$\langle\psi_{\text{phys}}|\partial^\mu A_\mu|\psi_{\text{phys}}\rangle = \langle\psi_{\text{phys}}|\partial^\mu A_\mu^+ + \partial^\mu A_\mu^-|\psi_{\text{phys}}\rangle, \tag{3.6.18}$$

$$= \langle\psi_{\text{phys}}|\partial^\mu A_\mu^+|\psi_{\text{phys}}\rangle + \langle\psi_{\text{phys}}|\partial^\mu A_\mu^-|\psi_{\text{phys}}\rangle = 0. \tag{3.6.19}$$

Clearly this equation can be separated into two conditions for physical states. It is enough to require

$$\begin{cases}\partial^\mu A_\mu^+|\psi_{\text{phys}}\rangle = 0,\\ \langle\psi_{\text{phys}}|\partial^\mu A_\mu^- = 0\end{cases}. \tag{3.6.20}$$

These equations are the conditions that determine the physical state $|\psi_{\text{phys}}\rangle$. Now consider a photon propagating in the positive z direction. The four vector momentum of such propagation is given by

$$k=\begin{pmatrix}\omega\\0\\0\\\omega\end{pmatrix}. \tag{3.6.21}$$

Hence the first of (3.6.20) with the previous choices of $\xi(\vec{k},\lambda)$ and k yields

$$-i\int\frac{d^3k}{(2\pi)^3}\frac{1}{2\omega}\Big(\omega\alpha(\vec{k},0)-\omega\alpha(\vec{k},3)\Big)e^{-ik\cdot x}|\psi_{\text{phys}}\rangle = 0, \tag{3.6.22}$$

$$\Rightarrow \Big(\alpha(\vec{k},0)-\alpha(\vec{k},3)\Big)|\psi_{\text{phys}}\rangle = 0. \tag{3.6.23}$$

Daggering this last equation, we immediately have

$$\langle\psi_{\text{phys}}|\Big(\alpha^\dagger(\vec{k},0)-\alpha^\dagger(\vec{k},3)\Big) = 0. \tag{3.6.24}$$

Introduce the number operators $\hat{N}_\mu$ (the $-$ sign on the right hand side of (3.6.26) is to ensure that $\hat{N}_0$ has positive eigenvalues)

$$\hat{N}_i = \int\frac{d^3k}{(2\pi)^3}\frac{1}{2\omega}\alpha^\dagger(\vec{k},i)\alpha(\vec{k},i), \tag{3.6.25}$$

$$\hat{N}_0 = -\int\frac{d^3k}{(2\pi)^3}\frac{1}{2\omega}\alpha^\dagger(\vec{k},0)\alpha(\vec{k},0). \tag{3.6.26}$$

If we combine the equations (3.6.22) and (3.6.23) in the correct way, it is straightforward to see that

$$\begin{aligned}&\langle\psi_{\text{phys}}|[\alpha^\dagger(\vec{k},0)+\alpha^\dagger(\vec{k},3)][\alpha(\vec{k},0)-\alpha(\vec{k},3)]|\psi_{\text{phys}}\rangle \quad + \\ &\langle\psi_{\text{phys}}|[\alpha^\dagger(\vec{k},0)-\alpha^\dagger(\vec{k},3)][\alpha(\vec{k},0)+\alpha(\vec{k},3)]|\psi_{\text{phys}}\rangle \\ &= 2\left(\langle\psi_{\text{phys}}|\left[\alpha^\dagger(\vec{k},0)\alpha(\vec{k},0)-\alpha^\dagger(\vec{k},3)\alpha(\vec{k},3)\right]|\psi_{\text{phys}}\rangle\right)=0.\end{aligned}$$

Thus, in the basis of Hilbert space spanned by the eigenvectors of the number operators $\hat{N}_\mu$, this equation implies

$$0=\int\frac{d^3k}{(2\pi)^3}\frac{1}{2\omega}\langle\psi_{\text{phys}}|\alpha^\dagger(\vec{k},0)\alpha(\vec{k},0)-\alpha^\dagger(\vec{k},3)\alpha(\vec{k},3)|\psi_{\text{phys}}\rangle, \tag{3.6.27}$$

$$=-(n_0+n_3)\langle\psi_{\text{phys}}|\psi_{\text{phys}}\rangle. \tag{3.6.28}$$

(3.6.28) requires the following. First, if $\langle\psi_{\text{phys}}|\psi_{\text{phys}}\rangle\neq 0$ the number operator eigenvalues n_0 and n_3 must be zero. Thus, these physical states only contain polarizations which are transverse to the direction of propagation of photons. The last possibility is that $n_0+n_3\neq 0$, in which case $|\psi_{\text{phys}}\rangle$ must be a null vector state, i.e. $\langle\psi_{\text{phys}}|\psi_{\text{phys}}\rangle=0$.

The presence of the null state as a physical state is a signature that the theory enjoys a gauge symmetry. The final expression for the Hamiltonian in terms of the number operator is

$$H=\int\frac{d^3k}{(2\pi)^3}\frac{\omega}{2\omega}\left(-\alpha^\dagger(\vec{k},0)\alpha(\vec{k},0)+\sum_{j=1}^{3}\alpha^\dagger(\vec{k},j)\alpha(\vec{k},j)\right). \tag{3.6.29}$$

Thus,

$$\langle\psi_{\text{phys}}|H|\psi_{\text{phys}}\rangle=\int\frac{d^3k}{(2\pi)^3}\frac{1}{2}\sum_{j=1}^{2}\langle\psi_{\text{phys}}|\alpha^\dagger(\vec{k},j)\alpha(\vec{k},j)|\psi_{\text{phys}}\rangle. \tag{3.6.30}$$

Clearly only transverse polarized photons contribute.

3.7 Path Integral Quantization and the Faddeev-Popov formula

In the first chapter we introduced the path integral method and used instantons to develop a semi-classical approximation that can be applied in quantum field theory. In this section we develop the path integral further, by explaining how to quantize theories with a local (gauge) symmetry. We will again see an advantage of path integral quantization: it is much easier to formulate than Dirac's constrained quantization or Gupta-Bleuler quantization. Consider the quantization of the free vector field:

$$Z[\mathcal{J}_\mu]=\int\mathcal{D}A_\mu e^{iS},\quad\text{where}\quad S=-\int d^4x\left(\tfrac{1}{4}F^{\mu\nu}F_{\mu\nu}-\mathcal{J}^\mu A_\mu\right). \tag{3.7.1}$$

$Z[\mathcal{J}_\mu]$ is known as the generating functional of $A_\mu(x)$ correlation functions. The integral is over all possible field configurations $A_\mu(x)$ since we are now dealing with a field theory.

The physical interpretation of $Z[\mathcal{J}_\mu]$ is that it is the transition amplitude between a starting field configuration and a final configuration. Note that (3.7.1) is a formal statement. There is no guarantee that the integral is convergent. We know that the physical states are gauge invariant i.e. that A_μ and A'_μ with

$$A'_\mu=A_\mu+\partial_\mu\chi, \tag{3.7.2}$$

both describe the same physical state. Therefore the direct calculation of the integral in (3.7.1) sums infinitely many gauge transformations of each physical configuration. That will surely lead us to a divergent factor in the value of the path integral. In general, direct computation of the path integral always lead us to either an inconsistency or a divergence, if the theory enjoys a gauge symmetry.

To avoid this problem there are two equivalent ways to proceed:

1. We only integrate over physical states. In this approach we will use a gauge fixing condition to select one physical state from the set of all gauge equivalent ones.
2. We integrate over all possible field configurations but divide by the volume of the gauge group.

We will develop the ideas we need with a simple zero dimensional toy model.

Example: Consider a theory with two fields x and y such that only the difference between x and y is physically meaningful.

$$P_{xy} \equiv x - y. \tag{3.7.3}$$

Suppose that the action is given by $S = \left(P_{xy}\right)^2 = (x-y)^2$. This theory is invariant under the gauge transformation

$$\begin{cases} x \to x + a, \\ y \to y + a \end{cases}.$$

A sample of the field configuration is given in the following figure.

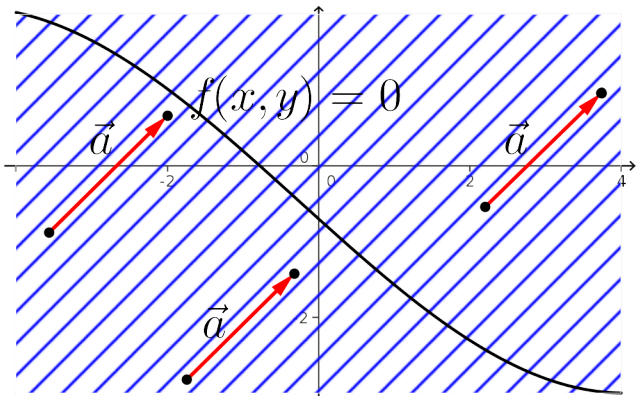


Figure 3.5: Field configuration and a gauge fixing condition. Each line of slope 1 is a physical state. Different points on a line are related by a gauge transformation. The dark curve is the line $f(x,y)=0$.

Now, the path integral is written

$$\mathcal{Z} = \int_{\mathbb{R}^2} dxdy e^{-(x-y)^2}. \tag{3.7.4}$$

Change variables from x, y to u, v with

$$\begin{aligned} u &= (x-y), \\ v &= \frac{x+y}{2}. \end{aligned}$$

the Jacobian J is defined by

$$dxdy = Jdudv = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial v}{\partial x} \\ \frac{\partial u}{\partial y} & \frac{\partial v}{\partial y} \end{vmatrix} dudv = dudv,$$

where the last equality follows by explicit calculation. Thus, the path integral (3.7.4) becomes

$$\mathcal{Z} = \int_{\mathbb{R}^2} dudve^{-u^2}, \tag{3.7.5}$$

$$= \sqrt{\pi} \int_{\mathbb{R}} dv. \tag{3.7.6}$$

Clearly, (3.7.6) tells us that $\mathcal{Z}$ is divergent. Indeed, it is not difficult to see that the divergent factor in (3.7.6) is the volume of the gauge group, which is the group of translations. Integrating over v is equivalent to integrating over all possible gauge equivalent copies of our physical states. Thus the divergent factor in (3.7.6) is expected. If we don't integrate over the extra copies we get $\mathcal{Z} = \sqrt{\pi}$ which we take as the correct value for (3.7.4).

We will now describe an elegant way of removing the volume of the gauge group. To this end, consider a gauge fixing condition $f(x, y) = 0$. The function f must have the following properties.

1. The line $f = 0$ must intersect every physical state once and only once.
2. Another condition is that the line $f = 0$ is not tangent to the line representing a physical state, since this selects the state more than once.

We illustrate in the figures below, examples of forbidden gauge fixing conditions $f(x, y) = 0$.

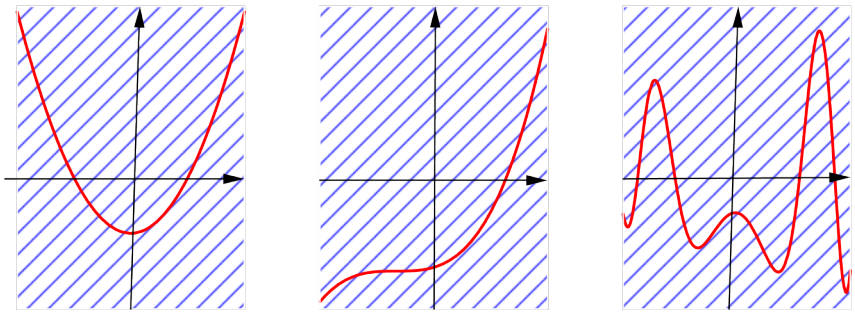


Figure 3.6: Forbidden gauge fixing conditions.

Now, for a gauge fixing condition f, insert the relation $\int df\delta(f) = 1$ in the integral defined in (3.7.4). We have

$$\mathcal{Z} = \int_{\mathbb{R}^2} dxdy \int df\delta(f)e^{-S}, \tag{3.7.7}$$

$$= \int_{\mathbb{R}^2} dxdy \int da \left|\frac{df}{da}\right|_{f=0} \delta(f)e^{-S}. \tag{3.7.8}$$

It is straightforward to see that $\delta(f)$ is independent of a. Also $\frac{df}{da}\Big|_{f=0}$ is independent of a. Indeed, the graph of $f = 0$ is unchanged by a gauge transformation since it defines what we mean by $a = 0$. Therefore (3.7.8) can be rewritten as

$$\mathcal{Z} = \int_{\mathbb{R}^2} dxdy \left|\frac{df}{da}\right|_{f=0} \delta(f)e^{-S} \int da \tag{3.7.9}$$

The last factor in the above integral is the volume of our gauge group. We can normalize $\mathcal{Z}$ by dividing it by the volume of the gauge group. Thus we have

$$\hat{\mathcal{Z}} = \int_{\mathbb{R}^2} dxdy \left|\frac{df}{da}\right|_{f=0} \delta(f)e^{-S}. \tag{3.7.10}$$

As a concrete example, for a gauge fixing condition use

$$\begin{aligned} &f : y = kx, \quad k \neq 1. \\ &\Rightarrow f \to f + a(1-k). \end{aligned}$$

We have

$$\frac{df}{da} = 1 - k.$$

Inserting this value into the above integral yields

$$\hat{\mathcal{Z}} = \int_{\mathbb{R}} dx \int_{\mathbb{R}} dy(1-k)\delta(y-kx)e^{-(x-y)^2}, \tag{3.7.11}$$

$$= \int_{\mathbb{R}} dx(1-k)e^{-(1-k)^2x^2} = \sqrt{\pi}. \tag{3.7.12}$$

We have shown that $\hat{\mathcal{Z}}$ is indeed the correct value for the integral (3.7.4) after the volume of the gauge group has been removed.

Similarly, it is straightforward to see that after removing the volume of the gauge group defined in (3.7.2) the generating functional in (3.7.1) becomes

$$\hat{\mathcal{Z}}_{\text{FP}} = \int \mathcal{D}A_\mu \det\left[\frac{\delta F[A_\mu(x)]}{\delta\chi(y)}\right]_{F=0} \prod \delta\big(F(A_\mu)\big)e^{iS}. \tag{3.7.13}$$

(3.7.13) is known as the Faddeev-Popov ansatz. $F(A_\mu) = 0$ is the gauge fixing condition, and the functional derivative is defined as follows

$$\frac{\delta\mathcal{F}[X(x)]}{\delta X(y)} \underset{\epsilon\to 0}{=} \frac{\mathcal{F}[X(x) + \epsilon\delta(x-y)] - \mathcal{F}[X(x)]}{\epsilon}. \tag{3.7.14}$$

Now, armed with this formula, the study of quantum electrodynamics in the Coulomb gauge can be carried out using the path integral formalism. Recall that the conditions for the Coulomb gauge are

$$\begin{aligned} F_1 &= A^0(\vec{x}, t) = 0, \\ F_2 &= \vec{\nabla} \cdot \vec{A}(\vec{x}, t) = 0. \end{aligned}$$

Hence we have

$$\frac{\delta F_1}{\delta\chi(y)} = \frac{\delta\left(A_0(\vec{x},t)+\partial_0\chi(x)\right)}{\delta\chi(y)} = \partial_0\delta^4(x-y), \tag{3.7.15}$$

$$\frac{\delta F_2}{\delta\chi(y)} = \frac{\delta\left(\vec{\nabla}\cdot\vec{A}(\vec{x},t)+\vec{\nabla}\cdot\vec{\nabla}\chi(x)\right)}{\delta\chi(y)} = \vec{\nabla}^2\delta^4(x-y) \tag{3.7.16}$$

We can now rewrite (3.7.13) using these equations. We find

$$\hat{\mathcal{Z}}_{\text{FP}} = \int \mathcal{D}A^0\mathcal{D}\vec{A}\left|\det\left(\partial^0\right)\right|\left|\det\left(\vec{\nabla}^2\right)\right|\prod\delta\big(F(A_\mu)\big)e^{iS}. \tag{3.7.17}$$

The Faddeev-Popov determinants $\det\left(\frac{\partial F_1}{\delta\chi}\right)$ and $\det\left(\frac{\partial F_2}{\delta\chi}\right)$ are independent of A_μ so that they can be absorbed into the measure of the path integral. The delta functions $\delta(F_1)$ and $\delta(F_2)$ set the temporal and longitudinal polarizations to zero. Thus, we only quantize the transverse photon polarization so that we have recovered our previous result from Dirac's constrained quantization.

3.8 Why do we need a "covariant derivative"?

The principle of general covariance requires that the laws of physics must not depend on our choice of coordinates. Note that this principle is very similar to the gauge principle. We write our equations using tensors to satisfy the principle of general covariance. The quantity $\partial_\mu V^\rho$, where V^ρ is a vector, is not a tensor. Indeed, under the coordinate transformation $x^\mu \to \tilde{x}^\mu$, we have

$$\begin{aligned}
\tilde{V}^\sigma &= \frac{\partial\tilde{x}^\sigma}{\partial x^\rho}V^\rho.\\
\Rightarrow \frac{\partial}{\partial\tilde{x}^\nu}\tilde{V}^\nu &= \frac{\partial x^\mu}{\partial\tilde{x}^\nu}\frac{\partial}{\partial x^\mu}\left(\frac{\partial\tilde{x}^\nu}{\partial x^\rho}V^\rho\right),\\
&= \frac{\partial x^\mu}{\partial\tilde{x}^\nu}\frac{\partial\tilde{x}^\nu}{\partial x^\rho}\frac{\partial}{\partial x^\mu}V^\rho + \frac{\partial x^\mu}{\partial\tilde{x}^\nu}\frac{\partial^2\tilde{x}^\nu}{\partial x^\mu\partial x^\rho}V^\rho.
\end{aligned}$$

This last equation tells us that the quantity $\partial_\mu V^\rho$ doesn't transform like a tensor under a change of coordinate. It is the last term above that spoils the transformation. The derivative of a general tensor of higher order is also not a tensor. Hence, we must change how we differentiate tensors in order to obtain a new tensor. This concept of differentiation is called **covariant differentiation**. This new concept is required to satisfy the principle of general covariance. The ideas developed in this section will be crucial to develop non-Abelian gauge theory, as we will explain below.

3.8.1 Lie derivative. In general relativity one possible covariant derivative is given by the Lie derivative of a vector field. The Lie derivative of a vector field A^α along a curve with tangent vector u^β is defined by

$$\mathcal{L}_u A^\alpha = u^\beta\partial_\beta A^\alpha - A^\beta\partial_\beta u^\alpha. \tag{3.8.1}$$

The Lie derivative of a vector (tensor) field is an extension of the derivative of a vector (tensor) field defined in flat space (or spacetime) to a curved space (or spacetime). Indeed, the flat space formula needs to be revised because on a general curved manifold, vectors (tensors) defined at different points do not belong to the same vector space.

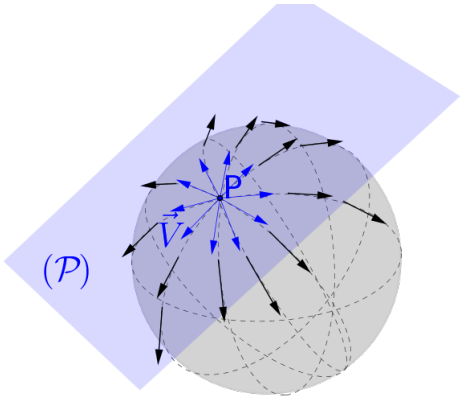


Figure 3.7: The fluid shown is flowing on the surface of the sphere.

This figure is an illustration of the flow of fluid directed perpendicularly from a certain point P on a sphere. The dashed line represents the streamlines followed by particles of the fluid. The plane $(\mathcal{P})$ is the tangent plane at the point P, where the flow starts. The vector $\vec{V}$ can be thought of as the velocity of a particle of the fluid. Consider the situation shown in Figure 3.7. If we follow the motion of a particle of the fluid along its streamline, we can precisely calculate the rate change of the velocity vector $\vec{V}$, along this streamline. The Lie derivative defined in (3.8.1) gives exactly this rate change of $\vec{V}$. In general if $\vec{u}$ is a tangent vector along a line defined on a manifold, the Lie derivative of $\vec{V}$ along that line is given by

$$\mathcal{L}_{\vec{u}}\vec{V} = u^i \partial_i V^j - V^i \partial_i u^j. \tag{3.8.2}$$

We now develop the mathematical construction of the Lie derivative. Consider a vector field denoted by $A^\alpha(x)$ and a set of curves parameterized by τ, σ, as illustrated in Figure 3.8. Note that this time we work in a 4-dimensional curved spacetime manifold in order to obtain a Lorentzian form of the Lie derivative. The mathematical definition of the Lie derivative of a vector field A^μ along the curve of tangent $u^\alpha = \frac{\partial x^\alpha}{\partial \tau}\big|_\sigma$ at a point P is

$$\mathcal{L}_u A^\alpha = \lim_{d\tau \to 0} \frac{A^\alpha(Q) - A^\alpha_L(P)}{d\tau}. \tag{3.8.3}$$

$A^\alpha_L(P)$ is the vector image of $A^\alpha(P)$ transported by a differential map defined along the curve $\mathcal{C}_{\sigma_1}$ of constant σ to the point Q as illustrated in Figure 3.8. The curves $\mathcal{C}_{\sigma_i}$ are members of a family of curves that are given. Their tangents vectors are u^α. $\mathcal{S}_{\tau_1}$ is defined so that its tangent at P is $A^\mu(P)$. the curve $\mathcal{S}_{\tau_1}$ passes through one point on each curve $\mathcal{C}_{\sigma_i}$. Call this point $\tau_1 = 0$. Start from $\tau_1 = 0$ on each curve and flow to $\tau_2 = d\tau$ using the tangent vectors u^α. Join τ_2 on each curve to get $\mathcal{S}_{\tau_2}$. The tangent vector to $\mathcal{S}_{\tau_2}$ at $\tau_2 = d\tau$ (this is the point Q) is $A^\mu_L(P)$. This is how we use the family of curves to move A^μ from P to Q.

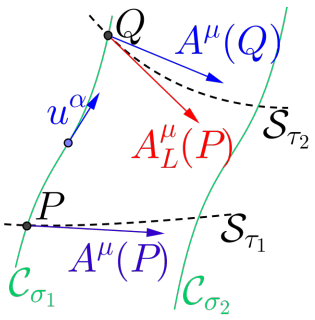


Figure 3.8: Illustration of the Lie derivative. The curve $\mathcal{S}_\tau$, links points on the curves $\mathcal{C}_{\sigma_i}$ with $\tau = 0$. The curve $\mathcal{S}_{\tau_2}$ links points on the curves $\mathcal{C}_{\sigma_i}$ with $\tau = d\tau$. $\mathcal{S}_{\tau_1}$ is constructed by declaring that $A^\mu(P)$ is tangent to $\mathcal{S}_{\tau_1}$. $\mathcal{S}_{\tau_2}$ is constructed by using the tangents u^α to flow to $d\tau$ from 0.

Points along each curve are labeled by τ. Each curve is labeled by a value of σ. Any point in spacetime that we consider is at a specific point on a specific curve, i.e. we can write $x^\mu(\sigma, \tau)$. We have two tangent vectors

$$u^\alpha = \left.\frac{\partial x^\alpha}{\partial \tau}\right|_\sigma \qquad n^\alpha = \left.\frac{\partial x^\alpha}{\partial \sigma}\right|_\tau$$

At P

$$A^\alpha(P) = n^\alpha(P) = \left.\frac{\partial x^\alpha}{\partial \sigma}\right|_P .$$

At Q

$$A^\alpha_L(P) = n^\alpha(Q) = \left.\frac{\partial x^\alpha}{\partial \sigma}\right|_Q .$$

Note that $A^\alpha_L(P)$ is located at Q. Now,

$$A^\mu(Q) = A^\mu(P) + u^\nu \partial_\nu A^\mu|_P \, d\tau, \tag{3.8.4}$$
$$A^\mu_L(P) = A^\mu(P) + A^\nu \partial_\nu u^\mu|_P \, d\tau. \tag{3.8.5}$$

To obtain (3.8.5) note that

$$\begin{aligned}
\frac{\partial A^\mu(P)}{\partial x^\nu} u^\nu &= \frac{\partial A^\mu(P)}{\partial x^\nu}\frac{\partial x^\nu}{\partial \tau} = \frac{\partial A^\mu(P)}{\partial \tau}, \\
&= \frac{\partial n^\mu(P)}{\partial \tau} = \left.\frac{\partial^2 x^\mu}{\partial \tau \partial \sigma}\right|_P , \\
&= \left.\frac{\partial^2 x^\mu}{\partial \sigma \partial \tau}\right|_P = \frac{\partial}{\partial \sigma} u^\mu(P), \\
&= \left.\frac{\partial x^\nu}{\partial \sigma}\frac{\partial u^\mu}{\partial x^\nu}\right|_P = A^\nu \left.\frac{\partial u^\mu}{\partial x^\nu}\right|_P .
\end{aligned}$$

Hence, subtracting (3.8.4) from (3.8.5) and using (3.8.3) yields

$$\mathcal{L}_u A^\mu = u^\nu \partial_\nu A^\mu - A^\nu \partial_\nu u^\mu. \tag{3.8.6}$$

3.8.2 Covariant Derivative from a Connection. A second way to obtain a covariant derivative, is by using a connection $\Gamma^{\mu}_{\nu\rho}$. The covariant derivative of a vector V^{ρ} on a curved spacetime is defined by

$$\nabla_{\mu}V^{\rho} = \partial_{\mu}V^{\rho} + \Gamma^{\rho}_{\mu\lambda}V^{\lambda}, \tag{3.8.7}$$

where $\Gamma^{\rho}_{\mu\lambda}$ is known as the Christoffel symbol. It is symmetric under the swap of the two subscripts μ and λ. The Christoffel symbol is not a tensor, but its presence in the covariant derivative is required so that the quantity $\nabla_{\mu}V^{\rho}$ is a tensor. It is now extremely simple to show that the Lie derivative is covariant

$$A^{\mu}\nabla_{\mu}V^{\rho} - V^{\mu}\nabla_{\mu}A^{\rho} = A^{\mu}\left(\partial_{\mu}V^{\rho} + \Gamma^{\rho}_{\mu\lambda}V^{\lambda}\right) - V^{\mu}\left(\partial_{\mu}A^{\rho} + \Gamma^{\rho}_{\mu\lambda}A^{\lambda}\right), \tag{3.8.8}$$

$$= A^{\mu}\partial_{\mu}V^{\rho} - V^{\mu}\partial_{\mu}A^{\rho}. \tag{3.8.9}$$

Given a metric $g^{\mu\nu}$ on a curved spacetime, if the covariant derivative $\nabla_{\rho}g^{\mu\nu}$ is equal to zero we say that the connection is metric compatible. This condition ensures, for example, that the inner product of two non null vectors is invariant after parallel transporting them. The Christoffel symbol is metric compatible. The Christoffel symbol can be related to the metric tensor as follows

$$\Gamma^{\rho}_{\mu\nu} = \frac{1}{2}g^{\rho\sigma}(\partial_{\mu}g_{\sigma\nu} + \partial_{\nu}g_{\mu\sigma} - \partial_{\sigma}g_{\mu\nu}). \tag{3.8.10}$$

3.8.3 Interpretation of $\Gamma^{\rho}_{\mu\nu}$. Consider the two dimensional Euclidean space $\mathbb{R}^2$. The metric of this space, using the polar coordinates, is given by

$$ds^2 = dr^2 + r^2 d\theta^2. \tag{3.8.11}$$

The matrices representing the tensor metric g_{ab}, $a, b \in \{r, \theta\}$ and its inverse g^{ab} are respectively

$$g_{ab} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}_{ab}, \quad g^{ab} = \begin{pmatrix} 1 & 0 \\ 0 & r^{-2} \end{pmatrix}_{ab}.$$

Using the formula (3.8.10), the relevant components of the Christoffel symbol are those where the partial derivative with respect to r is not zero. Hence, we have

$$\Gamma^{r}_{\theta\theta} = -r, \qquad \Gamma^{\theta}_{\theta r} = \Gamma^{\theta}_{r\theta} = \frac{1}{r}, \tag{3.8.12}$$

$$\Gamma^{\theta}_{\theta\theta} = \Gamma^{r}_{rr} = 0, \qquad \Gamma^{\theta}_{\theta r} = \Gamma^{\theta}_{r\theta} = 0. \tag{3.8.13}$$

From the equation (3.8.11) we can define two vectors such that

$$\begin{cases} \vec{e}_r \cdot \vec{e}_r = 1 & \Rightarrow \vec{e}_r = \hat{r} \\ \vec{e}_\theta \cdot \vec{e}_\theta = r^2 & \Rightarrow \vec{e}_\theta = r\hat{\theta}. \end{cases}$$

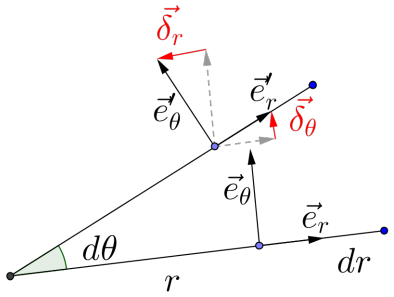


Figure 3.9: The change in the vectors $\vec{e_\theta}$ and $\vec{e_r}$ under a small displacement $d\theta$.

The small displacement vectors $\vec{\delta_a}$ in the above figure directly determines the coefficient Γ^a_{bc} calculated in (3.8.12). Indeed, from the figure we read

$$\vec{e_\theta}' - \vec{e_\theta} = \vec{\delta_r} = -rd\theta\hat{r} = -rd\theta\vec{e_r} = d\theta\Gamma^r_{\theta\theta}\vec{e_r},$$
$$\vec{e_r}' - \vec{e_r} = \vec{\delta_\theta} = d\theta\hat{\theta} = d\theta\frac{\vec{e_\theta}}{r} = d\theta\Gamma^\theta_{r\theta}\vec{e_\theta}.$$

These equations tell us that the Christoffel symbols relate the vector spaces defined at different points on the manifold. Indeed, for the above example, the map is explicitly defined by the vectors translation $\vec{\delta_r}$ and $\vec{\delta_\theta}$. If we now compare the definition of the covariant derivative and the above equations it is not difficult to see that

$$\nabla_\theta\vec{e_\theta} = \Gamma^r_{\theta\theta}\vec{e_r} = -r\vec{e_r} \qquad \nabla_r\vec{e_\theta} = \Gamma^\theta_{r\theta}\vec{e_\theta} = \frac{1}{r}\vec{e_\theta},$$
$$\nabla_r\vec{e_r} = \Gamma^r_{rr}\vec{e_r} = 0 \qquad \nabla_\theta\vec{e_r} = \Gamma^\theta_{\theta r}\vec{e_\theta} = \frac{1}{r}\vec{e_\theta}.$$

3.9 Covariant derivative in non-Abelian gauge theory

Consider a collection of massive complex scalar fields $\phi^a(x)|_{a=1,\cdots,N}$ described by a Lagrangian density given by

$$\mathcal{L} = \partial_\mu\phi^{a\star}\partial^\mu\phi^a - m^2\phi^{a\star}\phi^a; \tag{3.9.1}$$
$$= \partial_\mu\vec{\phi}^\dagger\partial^\mu\vec{\phi} - m^2\vec{\phi}^\dagger\vec{\phi}. \tag{3.9.2}$$

Clearly this theory is invariant under the global $U(N)$ group, the unitary group of $N \times N$ matrices. Indeed, for a transformation $U \in U(N)$ we have $U^\dagger U = \mathbb{1}$ and the transformation of the field is defined by

$$\begin{cases} \vec{\phi} \to U\vec{\phi} \\ \vec{\phi}^\dagger \to \vec{\phi}^\dagger U^\dagger \end{cases}.$$

Thus, the Lagrangian becomes

$$\mathcal{L}' = \partial_\mu(\vec{\phi}^\dagger U^\dagger)\partial^\mu(U\vec{\phi}) - m^2\vec{\phi}^\dagger U^\dagger U\vec{\phi},$$
$$= \partial_\mu\vec{\phi}^\dagger\partial^\mu\vec{\phi} - m^2\vec{\phi}^\dagger\vec{\phi},$$
$$\Rightarrow \mathcal{L}' = \mathcal{L}.$$

Above we used the fact that the transformation U is a global transformation so ∂_μ does not act on U.

We now derive the conserved Noether current associated to this global symmetry. Consider an element $U = e^{i\alpha^j(x)T^j}$ of the group, where $\alpha^j(x)$ and T^j are respectively the real transformation parameters and the generators of the group. The set $\{T^j\}$ is known as the basis of the vector space of the Lie algebra $\mathfrak{u}(N)$. Note that we allow the parameter α^j to depend on the spacetime coordinate x even though we study a global transformation. This is necessary for us to obtain the current density associated to the global symmetry. The dimension of the Lie algebra is N^2. Indeed, expanding the condition $U^\dagger U = \mathbb{1}$ up to the first order in the parameter $\alpha^j(x)$ yields

$$(\mathbb{1} - i\alpha^j T^{j\dagger})(\mathbb{1} + i\alpha^j(x)T^j) = \mathbb{1} - i\alpha^j T^{j\dagger} + i\alpha^j T^j = \mathbb{1},$$
$$\Rightarrow T^j = T^{j\dagger}.$$

This last equation implies that all the T^j are Hermitian. There are precisely $N(N-1)$ real degrees of freedom in the strict upper-half plus the N real degrees of freedom on the diagonal giving a total of N^2 different generators T^j of the group. Hence the dimension of the Lie algebra is N^2.

An infinitesimal gauge transformation of the field is now written as

$$\vec{\phi} \to (\mathbb{1} + i\alpha^j T^j + O(|\alpha|^2))\vec{\phi} \qquad \vec{\phi}^\dagger \to \vec{\phi}^\dagger(\mathbb{1} - i\alpha^j T^j + O(|\alpha|^2)),$$
$$\partial_\mu\vec{\phi} \to i\partial_\mu\alpha^j T^j\vec{\phi} + (\mathbb{1} + i\alpha^j T^j)\partial_\mu\vec{\phi} + O(|\alpha|^2) \quad \partial_\mu\vec{\phi}^\dagger \to \partial_\mu\vec{\phi}^\dagger - i\partial_\mu\vec{\phi}^\dagger\alpha^j T^j - i\vec{\phi}^\dagger\partial_\mu\alpha^j T^j + O(|\alpha|^2).$$

Under these transformation the change in the Lagrangian density, up to the first order in α^j, is given by

$$\mathcal{L} \to \mathcal{L}' = \partial_\mu\left[\vec{\phi}^\dagger(1 - i\alpha^j T^j)\right]\partial^\mu\left[\vec{\phi}(1 + i\alpha^j T^j)\right] - m^2\left[\vec{\phi}^\dagger(1 - i\alpha^j T^j)\right]\left[\vec{\phi}(1 + i\alpha^j T^j)\right], \tag{3.9.3}$$
$$= \partial_\mu\vec{\phi}^\dagger\partial^\mu\vec{\phi} - m^2\vec{\phi}^\dagger\vec{\phi} + i\partial_\mu\alpha^j\left(\partial^\mu\vec{\phi}^\dagger T^j\vec{\phi} - \vec{\phi}^\dagger T^j\partial^\mu\vec{\phi}\right). \tag{3.9.4}$$

The last term $J^k_\mu = \partial_\mu\vec{\phi}^\dagger T^k\vec{\phi} - \vec{\phi}^\dagger T^k\partial_\mu\vec{\phi}$ is the current density associated to the global symmetry $U(N)$. Indeed, this must be true since the total action of the theory must remain the same to first order when perturbed around a solution to the equation of motion. In fact, for a global symmetry we must have (this follows because the change δS must be first order in α and must vanish for constant α)

$$\delta S = \int d^4x J^k_\mu\partial^\mu\alpha^k \Rightarrow \delta S = -\int d^4x\partial^\mu J^k_\mu\alpha^k.$$

Around any solution to the equation of motion we have

$$\delta S = \int d^4x\partial^\mu J^k_\mu\alpha^k = 0 \Rightarrow \begin{cases}\partial^\mu J^k_\mu = 0\\ \forall\alpha^k \neq 0.\end{cases}$$

Now we want to build a theory that is invariant under a local $U(N)$ (gauge) symmetry. This time we must include the notion of the covariant derivative, since $\partial_\mu\vec{\phi}$ doesn't transform to $U\partial_\mu\vec{\phi}$. In the spacetime manifold, the covariant derivative of a vector is expressed as $\nabla_\mu V^\nu = (\delta^{\nu\lambda}\partial_\mu + \Gamma^\nu_{\mu\lambda})V^\lambda$. By analogy with ∇_μ, a covariant derivative in the space spanned by the field $\phi^a(x)$ has the form $D_\mu = \partial_\mu - ig\mathcal{A}_\mu$. We want to have the transformation law

$$\vec{\phi} \to \vec{\phi}' = U\vec{\phi}, \tag{3.9.5}$$
$$D_\mu\vec{\phi} \to (D_\mu\vec{\phi})' = U(D_\mu\vec{\phi}). \tag{3.9.6}$$

To find the transformation law that $\mathcal{A}_\mu$ satisfies, we need to manipulate the following identity

$$D'_\mu\vec{\phi}' \equiv (\partial_\mu - ig\mathcal{A}'_\mu)U\vec{\phi} = U(\partial_\mu - ig\mathcal{A}_\mu)\vec{\phi}. \tag{3.9.7}$$

Developing the left hand side of (3.9.7), we have

$$D'_\mu\vec{\phi}' = (\partial_\mu - ig\mathcal{A}'_\mu)U\vec{\phi}, \tag{3.9.8}$$
$$= (U\partial_\mu + \partial_\mu U - ig\mathcal{A}'_\mu U)\vec{\phi}, \tag{3.9.9}$$
$$= U(\partial_\mu + U^\dagger\partial_\mu U - igU^\dagger\mathcal{A}'_\mu U)\vec{\phi}. \tag{3.9.10}$$

Comparing (3.9.7) and (3.9.10) we identify

$$U^\dagger\partial_\mu U - igU^\dagger\mathcal{A}'_\mu U = -ig\mathcal{A}_\mu. \tag{3.9.11}$$

Using the identity $\partial_\mu(U^\dagger U) = (\partial_\mu U^\dagger)U + U^\dagger \partial_\mu U = 0$, equation (3.9.11) can be written as

$$\mathcal{A}'_\mu = U\mathcal{A}_\mu U^\dagger + \tfrac{i}{g} U\partial_\mu U^\dagger. \tag{3.9.12}$$

The transformation law $D_\mu \to D'_\mu$ can also be derived from $D'_\mu \vec{\phi}' = U D_\mu \vec{\phi}$. Indeed, we have

$$\begin{aligned} D'_\mu \vec{\phi}' &= D'_\mu U \vec{\phi}, && (3.9.13)\\ &= U D_\mu \vec{\phi} = U D_\mu U^\dagger (U\vec{\phi}); && (3.9.14)\\ \Rightarrow D'_\mu &= U D_\mu U^\dagger. && (3.9.15) \end{aligned}$$

Now define $g\mathcal{F}_{\mu\nu} \equiv i[D_\mu, D_\nu]$. By acting on an arbitrary $\vec{\phi}$ we have

$$\begin{aligned} i\left[(\partial_\mu - ig\mathcal{A}_\mu), (\partial_\nu - ig\mathcal{A}_\nu)\right]\vec{\phi} = &\left(i\partial_\mu\partial_\nu + g\partial_\mu\mathcal{A}_\nu + g\mathcal{A}_\nu\partial_\mu + g\mathcal{A}_\mu\partial_\nu - ig^2\mathcal{A}_\mu\mathcal{A}_\nu\right)\vec{\phi} - \\ &\left(i\partial_\nu\partial_\mu + g\partial_\nu\mathcal{A}_\mu + g\mathcal{A}_\mu\partial_\nu + g\mathcal{A}_\nu\partial_\mu - ig^2\mathcal{A}_\nu\mathcal{A}_\mu\right)\vec{\phi}. \end{aligned}$$

After simplifications the explicit form of $\mathcal{F}_{\mu\nu}$ is given by

$$\mathcal{F}_{\mu\nu} = \partial_\mu\mathcal{A}_\nu - \partial_\nu\mathcal{A}_\mu - ig[\mathcal{A}_\mu, \mathcal{A}_\nu]. \tag{3.9.16}$$

The transformation law $\mathcal{F}_{\mu\nu} \to \mathcal{F}'_{\mu\nu}$ can be obtained directly from the transformation law of $D_\mu \to D'_\mu$. Indeed, we have

$$g\mathcal{F}'_{\mu\nu} = i[UD_\mu U^\dagger, UD_\nu U^\dagger] = iU[D_\mu, D_\nu]U^\dagger = gU\mathcal{F}_{\mu\nu}U^\dagger. \tag{3.9.17}$$

Armed with the above results, a general form of the action for a theory invariant under both Lorentz and gauge transformations is

$$S = \int d^4x \left[(D_\mu\vec{\phi})^\dagger D^\mu\vec{\phi} - m^2\vec{\phi}^\dagger\vec{\phi} - \frac{1}{4}\mathrm{tr}(\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu})\right], \tag{3.9.18}$$

where the operator trace runs over the gauge group indices. The Lorentz invariance of this theory is ensured as usual by the proper contraction of Greek indices such as μ, ν, ... while the gauge invariance requires the trace. Indeed, using the cyclicity property of "tr", we have

$$\begin{aligned} \mathrm{tr}(\mathcal{F}'_{\mu\nu}\mathcal{F}^{\mu\nu\prime}) &= \mathrm{tr}\left(U\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu}U^\dagger\right) = \mathrm{tr}\left(U^\dagger U\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu}\right),\\ &= \mathrm{tr}(\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu}). \end{aligned}$$

In the next paragraph we will show that the fields $\mathcal{A}_\mu$ as well as $\mathcal{F}_{\mu\nu}$ are elements of the Lie algebra. To do this we need to work with the transformation law

$$\mathcal{A}'_\mu = U\mathcal{A}_\mu U^\dagger + \frac{i}{g}U\partial U^\dagger.$$

For an element $U = e^{i\alpha^a(x)T^a}$ where $T^a \in \mathfrak{u}(N)$, we have

$$\begin{aligned} U\partial_\mu U^\dagger &= e^{i\alpha^a(x)T^a}\partial_\mu e^{-i\alpha^a(x)T^a},\\ &= e^{i\alpha^a(x)T^a}e^{-i\alpha^a(x)T^a}(-i)\partial_\mu\alpha^a T^a = -i\partial_\mu\alpha^a T^a, \end{aligned}$$

which is Lie algebra valued. Hence $\mathcal{A}_\mu$ must take values in the Lie algebra. Since $\mathcal{F}_{\mu\nu}$ is constructed from $\mathcal{A}_\mu$, $\mathcal{F}_{\mu\nu}$ must also take values in the Lie algebra $\mathfrak{u}(N)$. Recall that the Lie algebra $\mathfrak{u}(N)$ of the

Lie group $U(N)$ is of dimension N^2. Thus, we define $\{T^j\}_{j=1,\ldots,N^2}$ to be the set of basis vectors of $\mathfrak{u}(N)$ such that

$$T^j = T^{j\dagger}\Big|_{j=1,\ldots,N^2}, \tag{3.9.19}$$

$$\mathrm{tr}(T^a T^b) = \delta^{ab}, \tag{3.9.20}$$

$$\sum_a T^{a\dagger}{}_{ij} T^a{}_{kl} = \delta_{ik}\delta_{jl} = \sum_a T^a{}_{ji} T^a{}_{kl}. \tag{3.9.21}$$

The first identity is required by the definition $UU^\dagger = 1$ while the second one is a normalization condition for the basis $\{T^j\}$. After normalizing the T^j, the last equation is nothing but the completeness relation of the Lie algebra. In addition to the above properties of T^j, we can also define the structure constants, f^{abc}, of the Lie algebra

$$[T^a, T^b] = i f^{abc} T^c. \tag{3.9.22}$$

It is straightforward to see that f^{abc} are real numbers. Indeed, by daggering this relation we have

$$\begin{aligned}
[T^a, T^b]^\dagger &= -i f^{abc\star} T^{c\dagger} = -i f^{abc\star} T^c, \\
&= -[T^a, T^b] = -i f^{abc} T^c; \\
\Rightarrow f^{abc\star} &= f^{abc}.
\end{aligned}$$

It is already clear that the f^{abc} are antisymmetric with respect to the first and second indices. It is not difficult to see that f^{abc} is completely antisymmetric. To see this, use the orthonormal condition in (3.9.20) and use the linearity plus cyclicity properties of the trace, to obtain

$$\begin{aligned}
\frac{1}{i}\mathrm{tr}(T^c[T^a, T^b]) &= f^{abd}\mathrm{tr}(T^c T^d) = f^{abc}, \\
&= \frac{1}{i}\mathrm{tr}(T^c T^a T^b - T^c T^b T^a) = \frac{1}{i}\mathrm{tr}(T^b T^c T^a - T^a T^c T^b), \\
&= \frac{1}{i}\mathrm{tr}([T^b, T^c] T^a).
\end{aligned}$$

These equations prove that f^{abc} is indeed completely antisymmetric. Now we can expand $\mathcal{A}_\mu$ and $\mathcal{F}_{\mu\nu}$ in the Lie algebra as

$$\mathcal{A}_\mu = A^a_\mu T^a, \tag{3.9.23}$$

$$\mathcal{F}_{\mu\nu} = F^a_{\mu\nu} T^a = \sum_a (\partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu) T^a. \tag{3.9.24}$$

It will be useful to work with the transformation law that $\mathcal{A}_\mu$ satisfies in terms of the components A^a_μ. Consider

$$\begin{aligned}
\mathcal{A}'_\mu &= U\mathcal{A}_\mu U^\dagger + \tfrac{i}{g} U\partial_\mu U^\dagger, \\
&= (\mathbb{1} + i\alpha^a T^a)\mathcal{A}_\mu(\mathbb{1} - i\alpha^a T^a) + \tfrac{i}{g}(\mathbb{1} + i\alpha^a T^a)(-i\partial_\mu \alpha^c) T^c, \\
&= A^c_\mu T^c + i\alpha^a [T^a, \mathcal{A}_\mu] + \tfrac{1}{g}\partial_\mu \alpha^c T^c, \\
\Rightarrow A^c_\mu{}' T^c &= \left(A^c_\mu - i\alpha^a A^b_\mu f^{abc} + \tfrac{1}{g}\partial_\mu \alpha^c\right) T^c.
\end{aligned}$$

This equation holds for all T^c so that

$$A^{a\prime}_\mu = A^a_\mu - f^{bca}\alpha^b A^c_\mu + \tfrac{1}{g}\partial_\mu \alpha^a, \tag{3.9.25}$$

$$= A^a_\mu - f^{abc}\alpha^b A^c_\mu + \tfrac{1}{g}\partial_\mu \alpha^a. \tag{3.9.26}$$

The field $F^a_{\mu\nu}$ is interpreted as the field strengths of N^2 interacting "photons". The interaction is a consequence of the commutator between the different fields $\mathcal{A}_\mu$. Hence the action found in (3.9.18) is not a free theory. This theory is closely related to Quantum Chromodynamics (QCD), in the case where the gauge group is $SU(3)$.

We have constructed a theory which is simultaneously Lorentz and gauge invariant. We are now interested in the quantization of the theory. We will again adopt the path integral quantization formalism. The gauge fixing condition that we are going to consider is the following

$$F^a[\mathcal{A}_\mu] = \partial_\mu A^{a\mu} = 0.$$

Note that these are the generalization of the Lorentz gauge fixing condition defined in the case where the gauge group is $U(1)$. The Faddeev-Popov determinant $\det\left(\frac{\delta F^a}{\delta \alpha^b}\right)$ is no longer constant but depends on $\mathcal{A}_\mu$. Indeed, with the help of (3.9.26) we have

$$\frac{\delta F^a[\mathcal{A}(x)]}{\delta \alpha^c(y)} = \partial_\mu \left[\left(g f^{abc} A^{b\mu} + \partial^\mu \delta^{ac}\right)\delta^4(x-y)\right].$$

To deal with the Faddeev-Popov determinant which now depends on the field $A^{b\mu}$, we introduce *ghost fields* $\mathcal{C}$ and $\bar{\mathcal{C}}$ which are complex Grassman numbers. Recall that the determinant of an $n \times n$ matrix $\mathcal{O}$ can be expressed as a Gaussian integral with the help of complex Grassman numbers as follows

$$\det(\mathcal{O}) = \int d\theta_1^\star d\theta_1 \dots d\theta_n^\star d\theta_n e^{-\sum_{i,j}\theta_i^\star \mathcal{O}_{ij}\theta_j}. \tag{3.9.27}$$

The Faddeev-Popov determinant can be computed in the same way. Hence we have

$$\det\left(\frac{\delta F^a}{\delta \alpha^c}\right) = \int \mathcal{D}\bar{\mathcal{C}}\mathcal{D}\mathcal{C} e^{iS_g}, \tag{3.9.28}$$

where $\mathcal{C}$, $\bar{\mathcal{C}}$ and S_g are defined by $\begin{cases} \mathcal{C} = C^a T^a \\ \bar{\mathcal{C}} = \bar{C}^a T^a \\ S_g = \int d^4x \bar{C}^a\left[-\partial^\mu\partial_\mu\delta^{ac} - g f^{abc}\partial^\mu A^b_\mu\right]C^c \end{cases}$

(3.9.28) is a path integral over the configurations of the ghost fields. The path integral of the theory, after removing the volume of the gauge group, takes the form

$$\mathcal{Z}_{\text{FP}}[\mathcal{J}_\mu] = \int \mathcal{D}\mathcal{A}_\nu \left|\det\left(\frac{\delta F^a}{\delta \alpha^c}\right)\right| \delta\left(\partial_\mu \mathcal{A}^\mu\right) e^{-i\int \text{tr}\left(\frac{1}{4g^2}\mathcal{F}_{\rho\sigma}\mathcal{F}^{\rho\sigma} + \mathcal{J}_\mu \mathcal{A}^\mu\right)}. \tag{3.9.29}$$

Using (3.9.28), this expression can also be written as

$$\mathcal{Z}_{\text{FP}}[\mathcal{J}_\mu] = \int \mathcal{D}\mathcal{A}_\nu \mathcal{D}\bar{\mathcal{C}}\mathcal{D}\mathcal{C}\delta\left(\partial_\mu \mathcal{A}^\mu\right) e^{iS + i\int dx^4 \text{tr}(\mathcal{J}_\mu \mathcal{A}^\mu)}, \tag{3.9.30}$$

where S is given by

$$S = \int d^4x \left\{\text{tr}\left(-\tfrac{1}{4g^2}\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu}\right) + \bar{C}^a\left[-\partial^\mu\partial_\mu\delta^{ac} - g f^{abc}\partial^\mu A^b_\mu\right]C^c\right\}. \tag{3.9.31}$$

In general it is much more convenient to use a Gaussian averaged gauge. To this end, consider an element $\xi \in \mathfrak{u}(N)$ and a gauge fixing condition such as

$$\partial_\mu \mathcal{A}^\mu - \xi = 0 \tag{3.9.32}$$

transform the action and path integral in the following manner

$$\mathcal{Z}_{\mathsf{FP}}[\mathcal{J}_\mu] = \int \mathcal{D}\mathcal{A}_\nu \mathcal{D}\bar{\mathcal{C}}\mathcal{D}\mathcal{C}\mathcal{D}\xi\delta\left(\partial_\mu \mathcal{A}^\mu - \xi\right) e^{iS + i\mathsf{tr}(\mathcal{J}_\mu \mathcal{A}^\mu)}, \quad \text{where}$$
$$S = \int d^4x \left\{ \mathsf{tr}\left(-\tfrac{1}{4g^2}\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu}\right) + \bar{C}^a\left[-\partial^\mu\partial_\mu\delta^{ac} - g f^{abc}\partial^\mu A^b_\mu\right]C^c + \tfrac{\alpha}{2}\mathsf{tr}(\xi^2)\right\}$$

The additional Gaussian integral over ξ simply contributes a constant factor that can be absorbed into the measure. Here again, by setting $\alpha = 1$ we will have the Feynman gauge fixing condition, familiar from what we have seen in the Gupta-Bleuler formalism. In fact we showed that the Feynman gauge adds extra terms such as $-\frac{1}{2}(\partial_\mu A^{\mu a})^2$ to the Lagrangian density. These terms are also evident in the Lagrangian density we obtain here

$$\mathcal{L} = -\frac{1}{4}(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu)(\partial^\mu A^{\nu a} - \partial^\nu A^{\mu a}) - \frac{1}{2}(\partial_\mu A^{\mu a})^2 - \bar{C}^a\partial_\mu\partial^\mu C^a$$
$$- g f^{abc}(\partial_\mu A^a_\nu)A^{\mu b}A^{\nu c} - \frac{1}{4}g^2 f^{abc}f^{ade}A^b_\mu A^c_\nu A^{\mu d}A^{\nu e} - g\bar{C}^a f^{abc}\partial^\mu(A^b_\mu C^c)$$

The final simplified version of the Lagrangian density using the Feynman gauge fixing condition is

$$\mathcal{L} = -\frac{1}{2}\partial_\nu A^a_\mu \partial^\nu A^{\mu a} - \bar{C}^a\partial_\mu\partial^\mu C^a - g f^{abc}(\partial_\mu A^a_\nu)A^{\mu b}A^{\nu c}$$
$$- \frac{1}{4}g^2 f^{abc}f^{ade}A^b_\mu A^c_\nu A^{\mu d}A^{\nu e} - g\bar{C}^a f^{abc}\partial^\mu(A^b_\mu C^c). \tag{3.9.33}$$

The parameters that appear in the Lagrangian are not directly the parameters we measure in the lab. It is useful to split the parameters appearing in the Lagrangian into a physical coupling plus an extra piece known as a counterterm. For a more complete discussion, the reader is referred to [5, p. 325]. If we now take account of all possible counterterms consisting of either relevant or marginal terms consistent with the symmetries of the theory, the total action is

$$S = \int d^4x \left(\mathcal{L}_{\mathsf{ren}} + \mathcal{L}_{\mathsf{c.t}}\right), \tag{3.9.34}$$

where

$$\mathcal{L}_{\mathsf{ren}} = -\frac{1}{2}\partial_\nu A^a_\mu \partial^\nu A^{\mu a} - \bar{C}^a\partial_\mu\partial^\mu C^a - g f^{abc}(\partial_\mu A^a_\nu)A^{\mu b}A^{\nu c}$$
$$- \frac{1}{4}g^2 f^{abc}f^{ade}A^b_\mu A^c_\nu A^{\mu d}A^{\nu e} - g\bar{C}^a f^{abc}\partial^\mu(A^b_\mu C^c); \tag{3.9.35}$$
$$\mathcal{L}_{\mathsf{c.t}} = -\frac{1}{2}\delta_1\partial_\nu A^a_\mu \partial^\nu A^{\mu a} - \delta_2^{(C)}\bar{C}^a\partial_\mu\partial^\mu C^a - g\delta_1^{(3g)} f^{abc}(\partial_\mu A^a_\nu)A^{\mu b}A^{\nu c}$$
$$- \frac{1}{4}\delta_1^{(4g)}g^2 f^{abc}f^{ade}A^b_\mu A^c_\nu A^{\mu d}A^{\nu e} - g\delta_1^{(\bar{C})}\bar{C}^a f^{abc}\partial^\mu(A^b_\mu C^c). \tag{3.9.36}$$

3.10 Feynman rules of the gauge theory

Having the above Lagrangian in hand, we can carry out the calculation of the Feynman rules. Recall that for any field theory the Feynman propagator is determined by the two point function of the field. Thus, our first task is to compute the two point function of the gauge fields $\mathcal{A}_\mu$ and of the ghost fields

$\mathcal{C}$. Since these fields do not mix at quadratic order we can compute their two point functions, separately. For the gauge fields, we have

$$\langle 0|T\left(A^a_\mu(x_1)A^b_\nu(x_2)\right)|0\rangle = \frac{\delta}{\delta J^a_\mu(x_1)}\frac{\delta}{\delta J^b_\nu(x_2)}\int \mathcal{D}A^d_\rho e^{iS_2+\int d^4x J^a_\mu A^{a\mu}}\Big|_{J^a_\mu=0}. \tag{3.10.1}$$

Above S_2 is the action of quadratic terms in the Lagrangian $\mathcal{L}_{\text{ren}}$ which are the only terms needed in the evaluation the two point function. The only quadratic gauge field term in (3.9.35) is the term $-\frac{1}{2}\partial_\nu A^a_\mu \partial^\nu A^{\mu a}$. Thus, (3.10.1) becomes

$$\langle 0|T\left(A^a_\mu(x_1)A^b_\nu(x_2)\right)|0\rangle = \frac{\delta}{\delta J^a_\mu(x_1)}\frac{\delta}{\delta J^b_\nu(x_2)}\int \mathcal{D}A^d_\rho e^{\int d^4x\left(-i\frac{1}{2}\partial_\mu A^a_\nu \partial^\mu A^{a\nu}+J^a_\mu A^{a\mu}\right)}\Big|_{J^a_\mu=0}, \tag{3.10.2}$$

$$= \frac{\delta}{\delta J^a_\mu(x_1)}\frac{\delta}{\delta J^b_\nu(x_2)}\int \mathcal{D}A^d_\rho e^{\int d^4x\left(i\frac{1}{2}A^a_\nu \partial_\mu \partial^\mu A^{a\nu}+J^a_\mu A^{a\mu}\right)}\Big|_{J^a_\mu=0}. \tag{3.10.3}$$

The passage to the last line has used an integration by parts. This is a Gaussian integral which is most easily performed by completing the square. Towards this end we set

$$i\partial_\mu\partial^\mu A^{a\nu}_o + J^{a\nu} = 0. \tag{3.10.4}$$

To solve this, introduce the Green's function $G(x-y)$, which obeys

$$\partial_\mu\partial^\mu G(x-y) = \delta(x-y). \tag{3.10.5}$$

Using the Green's function, to construct $A^{a\nu}_o$, shift

$$A^a_\mu(x) \longrightarrow A^{a\prime}_\mu = A^a_\mu + A^a_{o\mu}(x) = A^a{}_\mu(x) + i\int d^4y G(x-y)J^a(y). \tag{3.10.6}$$

Noting that the integral measure $\mathcal{D}A^d_\rho$ is invariant under the translation of the the gauge field in (3.10.6), (3.10.3) can be rewritten as

$$\langle 0|T\left(A^a{}_\mu(x_1)A^b_\nu(x_2)\right)|0\rangle = \frac{\delta}{\delta J^a_\mu(x_1)}\frac{\delta}{\delta J^b_\nu(x_2)}\int \mathcal{D}A^d_\rho e^{\int d^4x\left(i\frac{1}{2}A^{a\prime}{}_\nu \partial_\mu \partial^\mu A^{a\prime\nu}+J^{a\mu}A^{a\prime}{}_\mu\right)}\Big|_{J^a_\mu=0}. \tag{3.10.7}$$

We can simplify this expression if we substitute the explicit expression for the fields $\mathcal{A}'_\sigma$ into the argument of the exponential. Consider the argument of the exponential

$$\begin{aligned} I = \int d^4x \Bigg[i\frac{1}{2}\left(A^a_\nu(x) + i\int d^4y G(x-y)J^a_\nu(y)\right)\partial_\mu\partial^\mu\left(A^{a\nu}(x) + i\int d^4y G(x-y)J^{a\nu}(y)\right) \\ + J^a_\mu(x)\left(A^{a\mu}(x) + i\int d^4y G(x-y)J^{a\mu}(y)\right)\Bigg], \\ = \frac{i}{2}\int d^4x A^a_\nu(x)\partial_\mu\partial^\mu A^{a\nu}(x) - \frac{1}{2}\iint d^4x d^4y A^a_\nu(x)\partial_{x^\mu}\partial^{x^\mu}G(x-y)J^{a\nu}(y) \\ - \frac{1}{2}\iint d^4x d^4y G(x-y)J^a_\mu(y)\partial_{x^\nu}\partial^{x^\mu}A^{a\nu}(x) + i\iint d^4x d^4y J^a_\nu(x)G(x-y)J^{a\nu}(y) \\ - \frac{i}{2}\iint d^4x d^4y\left(\int d^4z \partial_{x^\mu}\partial^{x^\mu}G(x-z)J^{a\nu}(z)\right)G(x-y)J^a{}_\nu(y) + \int d^4x J^a_\nu(x)A^{a\nu}(x). \end{aligned}$$

Two successive integration by parts can be carried out to cancel the second, the third and the last terms of this equation. We need to use the equation (3.10.5) to finally obtain

$$I = \int d^4x \left(\frac{i}{2}\right) A^a_\nu(x)\partial_\mu\partial^\mu A^{a\nu}(x) + \frac{i}{2}\iint d^4x d^4y J^a_\nu(x) G(x-y) J^{a\nu}(y). \tag{3.10.8}$$

Thus, the expression for the two point function becomes

$$\langle 0|T\left(A^a_\mu(x_1)A^b_\nu(x_2)\right)|0\rangle = \frac{\delta}{\delta J^a_\mu(x_1)}\frac{\delta}{\delta J^b_\nu(x_2)}\int \mathcal{D}A^d_\rho e^{\frac{i}{2}\int d^4x A^a_\nu(x)\partial_\mu\partial^\mu A^{a\nu}(x) + \frac{i}{2}\iint d^4x d^4y J^a_\nu(x)G(x-y)J^{a\nu}(y)}\Big|_{J^a_\mu=0} \tag{3.10.9}$$

We can simplify this expression further by recalling the normalization

$$\int \mathcal{D}A^d_\rho e^{\frac{-i}{2}\int d^4x A^a_\nu(x)(-1)\partial_\mu\partial^\mu A^{a\nu}(x)} = 1,$$

so that the final expression for the two point function is

$$\langle 0|T\left(A^a_\mu(x_1)A^b_\nu(x_2)\right)|0\rangle = \frac{\delta}{\delta J^a_\mu(x_1)}\frac{\delta}{\delta J^b_\nu(x_2)}\exp\left(\frac{i}{2}\iint d^4x d^4y J^a_\nu(x)G(x-y)J^{a\nu}(y)\right)\Big|_{J^a_\mu=0}, \tag{3.10.10}$$

$$= i\delta^{ab}\eta_{\mu\nu}G(x_1-x_2). \tag{3.10.11}$$

This last equation, together with the equation (3.10.5), allows us to express the Feynman propagator, as

$$D^{ab}_{\mu\nu}(x_1,x_2) = \int \frac{d^4k}{(2\pi)^4}\frac{-ie^{ik\cdot(x_1-x_2)}}{k^2+i\epsilon}\eta_{\mu\nu}\delta^{ab}. \tag{3.10.12}$$

The corresponding Feynman diagram for this propagator, in momentum space, is

$$-i\frac{\eta_{\mu\nu}}{k^2+i\epsilon}\delta^{ab}: \quad a,\mu \xrightarrow{\ k\ } b,\nu.$$

Very similar calculations can be carried out in the evaluation of the Feynman propagator for the ghost fields. Indeed the two point function is

$$\langle 0|T\left(C^a(x_1)\bar{C}^b(x_2)\right)|0\rangle = \frac{\delta}{\delta\bar{\eta}^a(x_1)}\frac{\delta}{\delta\eta^b(x_2)}\int \mathcal{D}C^c\mathcal{D}\bar{C}^d e^{i\tilde{S}_2+\int d^4x(\bar{\eta}^aC^a+\eta^a\bar{C}^a)}\Big|_{\eta=\bar{\eta}=0}. \tag{3.10.13}$$

Again, $\tilde{S}_2$ is the action consisting of the quadratic terms in the ghost field appearing in $\mathcal{L}_{\text{ren}}$. The only quadratic term in the Lagrangian is the term $-\bar{C}^a\partial_\mu\partial^\mu C^a$. Recall that the ghost fields are Grassman numbers. Thus the source fields η^a and $\bar{\eta}^a$ have to be Grassman. The analog of (3.10.4) is

$$\begin{cases} -i\partial_\mu\partial^\mu C^a - \eta^a = 0, \\ i\partial_\mu\partial^\mu\bar{C}^a - \bar{\eta}^a = 0. \end{cases} \tag{3.10.14}$$

These equations share the same Green's function since the second equation is the complex conjugate of the first. Let $\tilde{G}(x-y)$ be the Green's function associated to (3.10.14). It satisfies

$$\partial_\mu\partial^\mu\tilde{G}(x-y) = \delta(x-y). \tag{3.10.15}$$

Accordingly, the shifted fields are expressed as

$$\begin{cases} C^{a\prime}(x) = C^a(x) + i\int d^4x\tilde{G}(x-y)\eta(y), \\ \bar{C}^{a\prime}(x) = \bar{C}^a(x) - i\int d^4x\tilde{G}(x-y)\bar{\eta}(y). \end{cases}$$

Now, use equation (3.10.15), two successive integration by parts and the anticommutativity of the Grassman numbers to obtain

$$\begin{aligned} I &= -i\int d^4x\left(\bar{C}^a(x) - i\int d^4y\tilde{G}(x-y)\bar{\eta}^a(y)\right)\partial_\mu\partial^\mu\left(C^a(x) + i\int d^4y\tilde{G}(x-y)\eta^a(y)\right) \\ &\quad + \int d^4x\left\{\eta^a(x)\left(\bar{C}^a(x) - i\int d^4x\tilde{G}(x-y)\bar{\eta}(y)\right) + \bar{\eta}^a(x)\left(\bar{C}^a(x) - i\int d^4x\tilde{G}(x-y)\bar{\eta}(y)\right)\right\}, \\ &= -i\int d^4x\bar{C}^a(x)\partial_\mu\partial^\mu C^a(x) + \int d^4x\bar{C}^a(x)\eta^a(x) - \int d^4x\bar{\eta}^a(x)C^a(x) - i\iint d^4xd^4y\tilde{G}(x-y)\bar{\eta}^a(y)\eta^a(x) \\ &\quad + \int d^4x\eta^a(x)\bar{C}^a(x) + \int d^4x\bar{\eta}^a(x)C^a(x) - i\iint d^4xd^4y\big(\eta^a(x)\tilde{G}(x-y)\bar{\eta}^a(y) - \bar{\eta}^a(x)\tilde{G}(x-y)\eta^a(y)\big), \\ &= -i\int d^4x\bar{C}^a(x)\partial_\mu\partial^\mu C^a(x) + i\iint d^4xd^4y\bar{\eta}^a(x)\tilde{G}(x-y)\eta^a(y). \end{aligned}$$

Again the first term of this last equation can be dropped because the path integral is normalized

$$\int \mathcal{D}C^c\mathcal{D}\bar{C}^d e^{-i\int d^4x\bar{C}^a(x)\partial_\mu\partial^\mu C^a(x)} = 1. \tag{3.10.16}$$

We now obtain

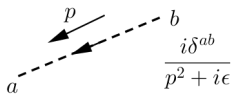
$$\langle 0|T\left(C^a(x_1)\bar{C}^b(x_2)\right)|0\rangle = \frac{\delta}{\delta\bar{\eta}^a(x_1)}\frac{\delta}{\delta\eta^b(x_2)}\exp\left(i\iint d^4xd^4y\bar{\eta}^a(x)\tilde{G}(x-y)\eta^a(y)\right)\Big|_{\eta=\bar{\eta}=0}, \tag{3.10.17}$$

$$= -i\delta^{ab}\tilde{G}(x_1-x_2). \tag{3.10.18}$$

The Feynman propagator of the ghost field is thus

$$\mathcal{D}^{ab}(x_1,x_2) = \int\frac{d^4p}{(2\pi)^4}\frac{ie^{ip\cdot(x_1-x_2)}}{p^2+i\epsilon}\delta^{ab}. \tag{3.10.19}$$

The corresponding Feynman diagram is



We now turn to the interaction terms in the Lagrangian $\mathcal{L}_{\text{ren}}$. The computation of the three point vertex of the gauge field is carried out by expanding the exponential of cubic terms in the Lagrangian and keeping the quadratic term in the exponential. The only cubic term in $\mathcal{L}_{\text{ren}}$ is the term $S_3 \propto -gf^{abc}\partial_\mu A^a_\nu A^b\mu A^{c\nu}$. Thus, at first order we have

$$D^{abc}_{\mu\nu\rho}(x_1,x_2,x_3) = \langle 0|T\left(A^a_\mu(x_1)A^b_\nu(x_2)A^c_\rho(x_3)\right)|0\rangle = \frac{\delta^3}{\delta J^{a\mu}(x_1)\delta J^{b\nu}(x_2)\delta J^{c\rho}(x_3)}\int\mathcal{D}A^d_\sigma S_3 e^{iS_2+\int d^4xJ^e_\mu A^{f\mu}}\Big|_{J^e_\mu=0}. \tag{3.10.20}$$

It is much easier to write this expression in terms of the possible Wick contractions. One example of the contractions of the fields is

$$D_1 = \int \mathcal{D}A^d_\sigma \int d^4z(-ig) f^{ijk} A^a_\mu(x_1)\partial_\tau A^i_\alpha(z) A^b_\nu(x_2) A^{j^\tau}(z) A^c_\rho(x_3) A^{k^\alpha}(z) e^{iS_2}.$$

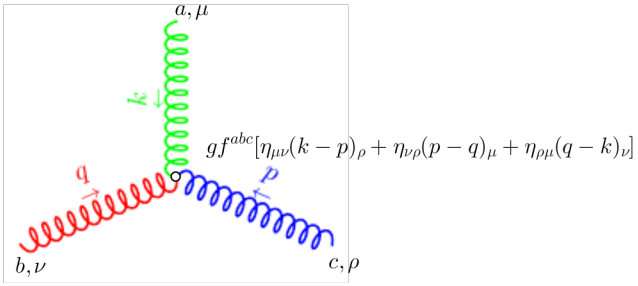
Now we can use the two point function for the gauge fields to obtain

$$D_1 = -igf^{ijk}\int d^4z(i)\left(\partial_{z^\tau} G(x_1-z)\right)\delta^{ai}\eta_{\mu\alpha}(i)G(x_2-z)\delta^{bj}\delta^\tau_\nu(i)G(x_3-z)\delta^{ck}\delta^\alpha_\rho, \tag{3.10.21}$$

$$= -igf^{abc}\int d^4z\left(\partial_{z^\nu}\int \frac{d^4k}{(2\pi)^4}\frac{ie^{ik\cdot(x_1-z)}}{k^2+i\epsilon}\right)\int\frac{d^4p}{(2\pi)^4}\frac{ie^{ip\cdot(x_2-z)}}{p^2+i\epsilon}\int\frac{d^4q}{(2\pi)^4}\frac{ie^{iq\cdot(x_3-z)}}{q^2+i\epsilon}, \tag{3.10.22}$$

$$= -gf^{abc}\eta_{\mu\rho}\iiint \frac{d^4k}{(2\pi)^4}\frac{d^4p}{(2\pi)^4}\frac{d^4q}{(2\pi)^4}k_\nu\frac{i}{k^2+i\epsilon}\frac{i}{p^2+i\epsilon}\frac{i}{q^2+i\epsilon}(2\pi)^4\delta(k+p+q)e^{i(k\cdot x_1+p\cdot x_2+q\cdot x_3)}. \tag{3.10.23}$$

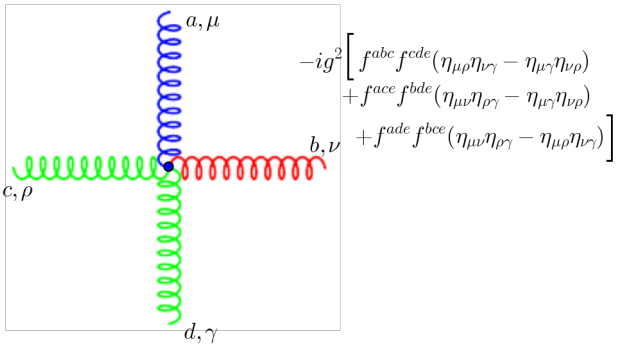
There are five possible contractions of the remaining fields. It is straightforward to see that the sum of these six contractions yields the following coefficient associated to the vertex in the Feynman diagram for the interaction of the gauge fields,



The total momentum flowing into the vertex is zero. The four point vertex of the gauge fields is evaluated by computing

$$D^{abcd}_{\mu\nu\rho\gamma}(x_1,x_2,x_3,x_4) = \frac{\delta^4}{\delta J^{a\,\mu}(x_1)\delta J^{b\,\nu}(x_2)\delta J^{c\,\rho}(x_3)\delta J^{d\,\gamma}(x_4)}\int \mathcal{D}A^l_\sigma S_4 e^{iS_2+\int d^4x J^e_\mu A^{f\,\mu}}\Big|_{J^e_\mu=0} \tag{3.10.24}$$

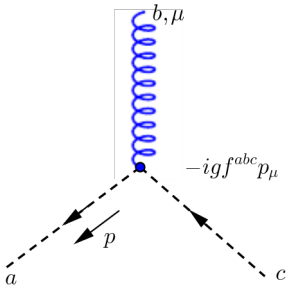
Noting that in $S_4 \propto -\frac{1}{4}g^2 f^{mnt}f^{mrs}A^n_{\mu_1}A^t_{\nu_1}A^{\mu_1\,r}A^{\nu_1\,s}$ there are two lower Lorentz indices and no derivative, it is clear that our result will involve a tensor of the form $\eta_{\mu\mu_1}\eta_{\nu\nu_1}\delta^{\mu_1}_\rho\delta^{\nu_1}_\gamma = \eta_{\mu\rho}\eta_{\nu\gamma}$. Indeed, the contributions of all the contractions of the gauge fields is summarized in the following diagram.



The total momentum flowing into the vertex is again zero. The last interaction term in the Lagrangian $\mathcal{L}_{\text{ren}}$ is $\tilde{S}_3 \propto -gf^{abc}\bar{C}^a\partial_\mu(A^{b\mu}C^c)$. The expression that we want to compute is

$$\langle 0|T\left(\bar{C}^a(x_1)A^b_\mu(x_2)C^c(x_3)\right)|0\rangle = \frac{\delta^3}{\delta\eta^a(x_1)\delta J^{b\mu}(x_2)\delta\bar{\eta}^c(x_3)}\int \mathcal{D}\bar{C}^i\mathcal{D}A^j_\rho\mathcal{D}C^k\tilde{S}_3 e^{iS_2+\int d^4y(J^e_\mu A^{e\mu}+\eta^e\bar{C}^e+\bar{\eta}^eC^e)}\Big|^{J^e_\mu=0}_{\eta^e=\bar{\eta}^e=0}. \tag{3.10.25}$$

Arguing as we did above we find



where the total momentum flowing into the vertex is again zero.

3.11 Renormalized perturbation theory

In this section we will develop the renormalized perturbation theory for our theory. We will develop the discussion in the context of an interacting scalar field. In this way, the ideas we need are developed in the simplest possible setting. Consider a field theory described by the Lagrangian

$$\mathcal{L} = \frac{1}{2}Z\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}Zm^2\phi^2 - gZ^2\phi^4.$$

Here g and m are two parameters appearing in the Lagrangian. We have no careful argument relating g to the interaction strength and m to the mass of the particle. Further, the field ϕ appearing in the Lagrangian might not represent the physical field strength that we measure. For that reason we have included a scaling Z which relates the field we actually measure to the field appearing in the Lagrangian. Renormalizing a theory amounts to eliminating Z, g and m for the physical mass m_p and the physical coupling g_{phys}.

Computations in QFT are plagued with divergences. We will see that renormalization will remove divergences, for a special class of theories know as renormalizable QFTs. Note however that this is a "side effect": the goal of renormalization is not to remove infinities, but to replace theory parameters with parameters we would measure in the laboratory. We will start with a discussion of the parameter Z. Towards this end, consider the free Klein-Gordon theory with the well known Lagrangian density

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2.$$

At the quantum level, we require that

$$[\phi(\vec{x},t),\dot{\phi}(\vec{y},t)] = i\delta(\vec{x}-\vec{y}) \Leftrightarrow [a(\vec{k}),a^\dagger(\vec{p})] = \delta(\vec{k}-\vec{p}).$$

Thus, a one particle state of this theory is

$$|\psi\rangle = \int d^3k f(\vec{k}) a^\dagger(\vec{k})|0\rangle.$$

To obtain a correctly normalized state, we require

$$\langle\psi|\psi\rangle = \int d^3k |f(\vec{k})|^2 = 1.$$

We usually interpret this last equation as the probability of creating a single particle state. For an interacting scalar field, the probability of creating single particle state is not equal to one. Indeed, since we consider an interacting theory, the scalar field creates not only a single particle piece but also multiparticle garbage. Thus, the field must be *renormalized* if we want it to create a single particle state with probability equal to 1. The scaling number Z implements this field renormalization. To see this, note that the Feynman propagator becomes

$$= \frac{iZ}{p^2 - m^2}.$$

This propagator is also obtained by quantizing

$$\mathcal{L} = Z^{-1}\left(\frac{1}{2}(\partial_\mu \phi')^2 - \frac{1}{2}m^2\phi'^2\right).$$

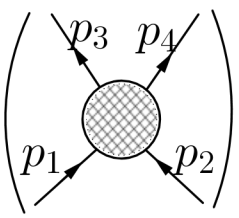
Canonical quantization of this Lagrangian gives

$$\begin{aligned}
[\phi'(\vec{x},t), \pi'(\vec{y},t)] = [\phi'(\vec{x},t), Z^{-1}\dot{\phi}'(\vec{y},t)] &= i\delta(\vec{x}-\vec{y}) \\
\Leftrightarrow [\phi'(\vec{x},t), \dot{\phi}'(\vec{y},t)] &= iZ\delta(\vec{x}-\vec{y}) \\
\Rightarrow [a'(\vec{k}), a'^\dagger(\vec{p})] &= Z\delta(\vec{k}-\vec{p}); \\
|\Psi\rangle = \int d^3k f(\vec{k}) a'^\dagger(\vec{k})|0\rangle \quad \langle\Psi|\Psi\rangle &= Z\int d^3k |f(\vec{k})|^2 \leq 1.
\end{aligned}$$

Now it is clear that our interacting field creates a single particle state with probability $Z \leq 1$.

The Lagrangian of the ϕ^4 theory given has three parameters $m,\ g,\ Z$. In 4 dimensions g is dimensionless. Renormalization, the process of eliminating Z, g, m for m_p, g_{phys}, is achieved through matching conditions (or renormalization conditions) given below.

- The pole of the two point function must be at m_p^2, where m_p is the physical mass that we measure during an experiment.
- The residue of the pole m_p must be equal to "i" to ensure that field creates single particle states with probability one. This condition sets the parameter Z.
- The prime in the following diagram means remove the external legs



$$' = ig_{\text{phys}}4! \ \text{ at } s = 4m_p^2 \ \ t = u = 0.$$

g_{phys} is defined to be the magnitude of the scattering amplitude at the specified momentum. The variables s, t and u are known as the Mandelstam variables. These variables are conventionally useful to express scattering amplitude in a simple way. In the above diagram they are defined by

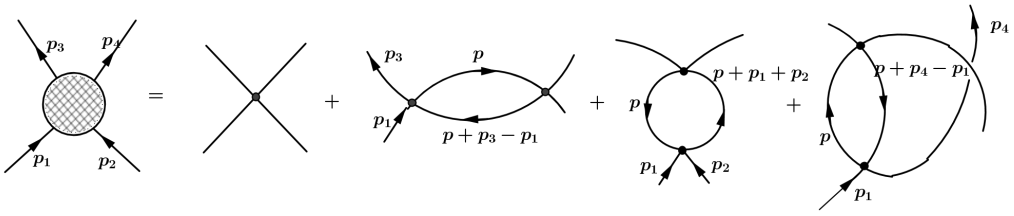
$$s = (p_1 + p_2)^2 = (p_3 + p_4)^2,$$
$$t = (p_3 - p_1)^2 = (p_4 - p_2)^2,$$
$$u = (p_4 - p_1)^2 = (p_3 - p_2)^2.$$

For more complete discussion, the reader is referred to [5, p. 156, 325-327].

Renormalization is always preceded by some regularization of the theory. A regularization of the theory is a deformation of the theory so that the original infinities only appear when the deformation parameter goes to zero. Some property of the theory is always spoiled by the regularization method. We summarize in the following table some methods of regularization and the properties of the theory they spoil.

Method	Summary	Spoils
Cut off	$p^\mu = (p^0, \vec{p}) \Rightarrow \lvert\vec{p}\rvert < \Lambda$	Scale invariance & Lorentz invariance
Pauli-Villars	add new mass M particle with " wrong statistics"	Scale invariance & Hermiticity
Dimension regularization	continue to 2ω dimensions	scale invariance

Note that the covariant cut off condition $p^2 = (p^0)^2 - |\vec{p}|^2 < \Lambda^2$ is ineffective as a regularization method. We will now consider a detailed study of our renormalization conditions. Start from the third condition above. The diagrams corresponding to the scattering amplitude $\mathcal{M}(p_1 p_2 \to p_3 p_4)$ at one loop are



Our renormalization condition is

$$-ig_{\text{phys}} = -igZ^2 + \frac{(ig)^2}{2}\int \frac{d^4p}{(2\pi)^4}\frac{i}{p^2-m^2}\left[\left.\frac{i}{(p+p_3-p_1)^2-m^2}\right|_{p_1=p_3} + \left.\frac{i}{(p+p_1+p_2)-m^2}\right|_{p_1+p_2=4m_p^2} + \left.\frac{i}{(p+p_4-p_1)^2-m^2}\right|_{p_1=p_4}\right] 4!Z^2 \tag{3.11.1}$$

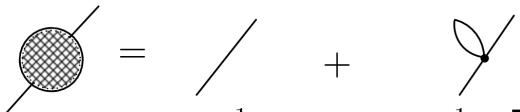
Introducing the quantity

$$iV(k^2) = \int \frac{d^4p}{(2\pi)^4}\frac{i}{p^2-m^2}\frac{i}{(p+k)^2-m^2}, \quad \text{where} \quad k = p_1 + p_2, \tag{3.11.2}$$

we can write our third renormalization condition as

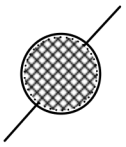
$$-ig_{\text{phys}} = -igZ^2 + \frac{(ig)^2}{2}[i\,V(s)|_{s=4m^2} + i\,V(t)|_{t=0} + i\,V(u)|_{u=0}]4!Z^2. \tag{3.11.3}$$

Consider now the first two renormalization conditions which both involve the two point function. At one loop we have



$$= \frac{iZ^{-1}}{p^2-m^2} + 12i\frac{Z^{-1}}{p^2-m^2}\left[\int \frac{d^4k}{(2\pi)^4}\frac{iZ^{-1}}{k^2-m^2}\right](igZ^2)\frac{iZ^{-1}}{p^2-m^2}.$$

For small coupling g and after some algebra, this can be written as



$$= \frac{iZ^{-1}}{p^2-m^2-12\int \frac{d^4k}{(2\pi)^4}\frac{ig}{k^2-m^2}}.$$

Hence, according to the renormalization condition above, the physical mass of the scalar field is

$$m_p^2 = m^2 - 12\int \frac{d^4k}{(2\pi)^4}\frac{ig}{k^2-m^2}. \tag{3.11.4}$$

The renormalization conditions also imply $Z = 1$.

There is a general method, known as power counting, that enables us to estimate what divergences we expect in a QFT. The method is basically a clever application of dimensional analysis. Indeed, in $d-$dimensional spacetime any Feynman diagram has the algebraic form

$$I = \int d^dp_1 \ldots \int d^dp_L \frac{i}{q_1^2-m^2}\cdots\frac{i}{q_I^2-m^2}, \tag{3.11.5}$$

where L is counting the number of loops while I is counting the number of internal propagators in the diagram. If we consider the contribution to these integrals coming from very large momenta, the masses may be set to zero. In this case the only dimensionfull parameter is the large momentum cut off Λ. Since this is the only dimensionfull parameter in the problem we can estimate how I depends on Λ using nothing but dimensional analysis.

In this way, it is straightforward to see that the convergence of the diagram is controlled by the quantity

$$D = dL - 2I. \tag{3.11.6}$$

Indeed, each loop in a Feynman diagram brings a potentially divergent d-momentum integral in the numerator. In contrast, each propagator aids the convergence of the integral by putting two powers of the momentum in the denominator. The quantity D is called the superficial degree of divergence because although it is a useful guide, it sometimes estimates the divergence incorrectly. Note that this argument must be modified when we include fermionic propagators, since a fermion propagator has a different momentum dependence

If $D < 0$ it can be shown that for a diagram without a divergent subdiagram, the integral (3.11.5) is convergent. In contrast, if $D = 0$ the divergence of (3.11.5) is logarithmic ($\log\Lambda$) and finally if $D > 0$,

the ultra-violet divergence is proportional to Λ^D. Now in general, the formula for the superficial degree of divergence is

$$D = d - \left(\frac{d}{2} - 1\right) E + n\left[\frac{r}{2}(d-2) - d\right], \tag{3.11.7}$$

where n counts the number of vertices in the diagram, E counts the external legs and r is the power of the field appearing in the interaction term as

$$\mathcal{L}_I = \int d^d x g_r \phi^r. \tag{3.11.8}$$

To derive (3.11.7) start by noting that the number of momenta we integrate over is given by the number of loops L. We also know that we need to integrate over a momentum for each internal line, that each vertex has a momentum conserving delta function collapsing one of the integrals except for one delta function which expresses the overall energy-momentum conservation of the diagram. Thus,

$$L = I - n + 1.$$

Next there are $E + 2I$ propagators ends that must connect to rn points on the vertices, so that

$$rn = E + 2I.$$

Using these two identities in (3.11.6) gives (3.11.7). From the interaction term, as in (3.11.8), we can argue that the dimension of the coupling g_r is the coefficient of the number n of vertices in the formula (3.11.7). In other words, we have

$$[g_r] = [\text{Length}]^{r\left(\frac{d-2}{2}\right)-d}.$$

Looking at (3.11.7) it is clear that any theory with a coupling of dimension $[g] = L^a$ with $a > 0$, has an infinite number of Feynman diagrams which diverge. These can't be removed by a finite number of renormalization conditions. Thus, these theories are not renormalizable.

Since the mass dimension is the inverse of the length dimension, the above statement can be rephrased as follows.

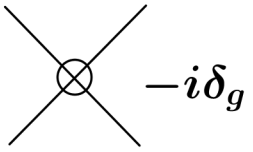
- A theory is renormalizable if the dimensions of all the coupling constants are either dimensionless or have positive mass dimension.
- A theory is non-renormalizable if the dimension of any of the coupling constants has a strictly negative power of the mass dimension.

The expressions relevant or marginal terms, refer respectively to couplings that have a positive power of the mass dimension or are dimensionless.

Now, returning to the ϕ^4 theory, the above argument implies that this theory is renormalizable in $d = 4$ dimensions. Recall that the Lagrangian density of this theory is

$$\begin{aligned}
\mathcal{L} &= \frac{1}{2} Z \partial_\mu \phi \partial^\mu \phi - \frac{1}{2} m^2 Z \phi^2 - Z^2 g \phi^4 \\
&= \frac{1}{2}(1 + Z - 1)\partial_\mu \phi \partial^\mu \phi - \frac{1}{2} m^2 (1 + Z - 1)\phi^2 - (1 + Z^2 - 1) g \phi^4 \\
&= \underbrace{\frac{1}{2}\partial_\mu \phi \partial^\mu \phi - \frac{1}{2} m^2 \phi^2 - g\phi^4}_{\text{renormalized terms}} + \underbrace{\delta_Z \partial_\mu \phi \partial^\mu \phi - \delta_{m^2} \phi^2 - \delta_g \phi^4}_{\text{counterterms}}.
\end{aligned}$$

The counterterms [5, p. 325], in the Lagrangian introduce new Feynman diagrams with coefficients chosen to ensure that we can satisfy the renormalization conditions order by order in perturbation theory. Thus, using the renormalization conditions we solve for the values of the counterterms order by oder by order in perturbation theory. As an example, the term $\delta_g \phi^4$ introduces the new diagram



4. Euclidean Yang-Mills theory

4.1 Finite action field configurations

One of our aims in this chapter is to study the Yang-Mills field configurations that have a finite Euclidean action. Recall that the solutions to the Euclidean equations of motion are called instantons. This chapter is a generalization of our study of instantons in quantum mechanics. From now on we will work in Euclidean space and any action that we use is an Euclidean action. The metric in Euclidean space is the Kronecker symbol $\delta_{\mu\nu}$. Naively the statement **finite action** i.e. $S < \infty$ seems to be required so that the path integral weight $e^{-S/\hbar}$ is non-zero. Indeed, after a Wick rotation of the time coordinate, the path integral of a scalar field

$$\mathcal{Z}[\mathcal{J}] = \int \mathcal{D}\phi e^{-\frac{S[\phi,\mathcal{J}]}{\hbar}} \tag{4.1.1}$$

seems to be dominated by the minimum of the action. Field configurations with $S = \infty$ do not seem to contribute to the path integral.

However, this is not true. As an illustration, go back to the quantum mechanical path integral of the harmonic oscillator, with Hamiltonian $H = \frac{1}{2}\left(\frac{dx}{dt}\right)^2 + \frac{\omega^2}{2}x^2$. We showed that a general path $x(t)$ can be expanded around any solution $\tilde{x}(t)$ as

$$\begin{aligned} x(t) &= \tilde{x}(t) + \sum_n C_n x_n(t), \\ \int_{-T/2}^{T/2} dt x_n x_m &= \delta_{nm}, \\ \left(-\frac{d^2}{dt^2} + \omega^2\right) x_n &= \lambda_n x_n. \end{aligned}$$

Evaluating the action, $S = \int dt H$ of this path, up to the second order in C_n, yields

$$S[x(t)] = \frac{1}{2}\sum_n b_n^2, \quad \text{where} \quad b_n = C_n\sqrt{\lambda_n}. \tag{4.1.2}$$

We normalize the path integral measure as

$$N'[dx(t)] = \prod_n \frac{db_n}{\sqrt{2\pi\hbar}}$$

so that

$$N'\int [dx]e^{-S/\hbar} = \prod_n \int_{-\infty}^{+\infty} \frac{db_n}{\sqrt{2\pi\hbar}} e^{-\frac{b_n^2}{2\hbar}} = 1 \tag{4.1.3}$$

Now, every finite action motion requires that $[b_n]$ lies inside some hypercube, $-L \leq b_n \leq L$. Under this condition it is straightforward to see that (4.1.3) becomes

$$N'\int [dx]e^{-S/\hbar} = \lim_{n\to\infty} \prod_n \underbrace{\left[\int_{-L}^{+L} \frac{db_n}{\sqrt{2\pi\hbar}} e^{-\frac{b_n^2}{2\hbar}}\right]}_{<1} \to 0. \tag{4.1.4}$$

Hence, leaving out the finite action motions does not affect the value of the path integral. The reason why we consider finite action configurations is because they lead to a non-zero answer when we use them as a starting point for a semi-classical expansion.

Consider the Yang-Mills action

$$S = \frac{1}{4\lambda^2}\int d^4x \mathrm{tr}(\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu}), \tag{4.1.5}$$

where λ is a coupling constant and we adopt the following conventions

$$\begin{aligned} D_\mu &= \partial_\mu + \lambda\mathcal{A}_\mu, \quad \mathcal{F}_{\mu\nu} = [D_\mu, D_\nu], \\ \Rightarrow \frac{1}{\lambda}\mathcal{F}_{\mu\nu} &= \partial_\mu\mathcal{A}_\nu - \partial_\nu\mathcal{A}_\mu + \lambda[\mathcal{A}_\mu, \mathcal{A}_\nu], \\ \mathcal{A}_\mu^\dagger &= -\mathcal{A}_\mu. \end{aligned}$$

With these conventions the fields are anti-Hermitian as are the generators of the gauge group. In 4-dimensional space, the integral measure d^4x is $d^4x = r^3 dr d\Omega_3$ where $r = \sqrt{\delta_{\mu\nu}x^\mu x^\nu}$ and Ω_3 are respectively the radius and the solid angle of a 3-dimensional sphere S^3. In general it is straightforward to show that the total solid angle of an $(n-1)$-dimensional sphere (S^{n-1}) is $\Omega_{n-1} = 2\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}$. Thus, for our case here we have $\Omega_3 = 2\pi^2$. Hence equation (4.1.5) becomes

$$S = \frac{1}{4\lambda^2}\int_0^{\Omega_3} d\Omega_3 \int_0^{+\infty} dr r^3 \mathrm{tr}(\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu}). \tag{4.1.6}$$

For a finite action $\mathrm{tr}(\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu})$ must fall off for large r at least as fast as $\frac{1}{r^{4+\epsilon}}\big|_{\epsilon>0}$. An equivalent statement of this condition is to say

$$\frac{1}{\lambda}\mathcal{F}_{\mu\nu} = \partial_\mu\mathcal{A}_\nu - \partial_\nu\mathcal{A}_\mu + \lambda[\mathcal{A}_\mu, A_\nu] \propto \frac{1}{r^{2+\epsilon}} \quad \Rightarrow \mathcal{A}_\mu \propto \frac{1}{r^{1+\epsilon}}. \tag{4.1.7}$$

This is a condition on the behaviour of the field for large r, that is on a three dimensional sphere S^3. As a consequence of gauge invariance, we can relax (4.1.7). Indeed, it is enough to require

$$\mathcal{A}_\mu = \frac{1}{\lambda}g\partial_\mu g^\dagger + O\left(\frac{1}{r^{1+\epsilon}}\right). \tag{4.1.8}$$

Above we have $g \in G$, where $G = U(N)$ is the gauge group. Clearly all elements of the gauge group G yielding a non zero field configuration on S^3_∞ must be maps from S^3_∞ to G. It is not difficult to see that the physical fields $\mathcal{F}_{\mu\nu}$ are always zero on S^3_∞. Even though our gauge field does not vanish as $r \longrightarrow \infty$, it still gives a finite action solution. We can summarize our discussion with the following corollary

Corollary: For every finite action field configuration on $\mathbb{R}^4$, there exists a map from S^3_∞ to the gauge group G defined by

$$g : S^3_\infty \longrightarrow G. \tag{4.1.9}$$

Let $\mathcal{S}_G$ be the set of maps from S^3_∞ to G, giving a finite action field configuration. Consider a non zero field configuration $\mathcal{A}_\mu|_{S^3_\infty} = \frac{1}{\lambda}g\partial_\mu g^\dagger$; it is not difficult to show that an element g, giving a configuration

of a finite action, is not unique. Towards this end, pick an element of the gauge group $h \in G$ and perform the gauge transformation

$$\mathcal{A}_\mu \to \mathcal{A}'_\mu = h\mathcal{A}_\mu h^\dagger + \frac{1}{\lambda} h\partial_\mu h^\dagger + O(\tfrac{1}{r^{1+\epsilon}}).$$

We can rewrite this as

$$\mathcal{A}'_\mu = h(\frac{1}{\lambda} g\partial_\mu g^{-1})h^{-1} + \frac{1}{\lambda} h\partial_\mu h^{-1} + O(\tfrac{1}{r^{1+\epsilon}}), \tag{4.1.10}$$

$$= \frac{1}{\lambda} hg(\partial_\mu g^{-1} h^{-1} + g^{-1}\partial_\mu h^{-1}) + O(\tfrac{1}{r^{1+\epsilon}}), \tag{4.1.11}$$

$$= \frac{1}{\lambda} hg\partial_\mu (hg)^{-1} + O(\tfrac{1}{r^{1+\epsilon}}), \tag{4.1.12}$$

$$= \frac{1}{\lambda} hg\partial_\mu (hg)^\dagger + O(\tfrac{1}{r^{1+\epsilon}}). \tag{4.1.13}$$

$\mathcal{A}'_\mu$ is the same physical state as $\mathcal{A}_\mu$. Note that in general we can not choose hg to be the identity of the gauge group. In fact the gauge transformation h is defined on $\mathbb{R}^4$ while g is only defined on S^3_∞. Thus, h must always be a smooth function and must not depend on the angles of the three sphere at radius $r = 0$. We can always choose $h = \mathbb{1}_G$ at $r = 0$.

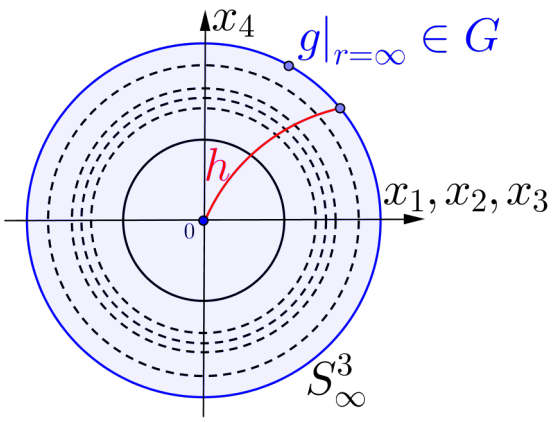


Figure 4.1: Figure illustrating how h and g are defined geometrically.

Accordingly if we can choose hg to be the identity, it follows that we must be able to continuously deform g to the identity. However we have no proof that we can do this. We will see later that we can't pick $h = g^{-1}$ in general. Note that the fact that we were working in 4-dimensional Euclidean space leads us to the manifold S^3 at $r = \infty$.

4.2 Some geometry and topology background.

Now we are interested in classifying the set of finite action field configurations. Recall that we denote the set of these configurations $\mathcal{S}_G$. It turns out that this classification can be performed by a simple set of rules. For any $g_1,\ g_2 \in \mathcal{S}_G$ if there exists $h \in G$ such that $g_1 = hg_2$ then the configurations corresponding to g_1 and g_2 are counted as the same physical configuration. Indeed, this must be true

since $h \in G$ is a gauge transformation implying that the configurations corresponding to g_1 and g_2 are the same physical state. Hence, two physical configurations are different if they can not be continuously deformed into each other by a gauge transformation, in other words $\nexists h \in G$ such that $g_1 = hg_2$. In order to carry out this classification, start with a simple toy model problem. Obviously the simplest version of this problem is to choose $G = U(1)$ and the Euclidean space $\mathbb{R}^2$. We start with some definitions.

- **Path.** Given a topological space X and an interval $I = [0, 1]$, a continuous map $\alpha : I \to X$ is called a path with initial point $\alpha(0) = x_0$ and final point $\alpha(1) = x_1$.
- **Loop.** A path is called a loop at the base point x_0 if $\alpha(0) = \alpha(1) = x_0$.
- **Constant path.** A constant path c_{x_0} is a surjective map from I to a single point $\{x_0\} \subset X$. Automatically, by definition, a constant path c_{x_0} is also a loop at the base point x_0.
- **Inverse path.** Given a path $\alpha(s)$ such that $\alpha(0) = x_0$ and $\alpha(1) = x_1$, the inverse path is defined by $\alpha^{-1}(s) = \alpha(1 - s)$. Thus, we have $\alpha^{-1}(0) = x_1$ and $\alpha^{-1}(1) = x_0$.
- **Product of two paths.** The product of two paths α and β is a path denoted $\alpha * \beta$ and can be defined as follows.

$$\alpha * \beta(s) = \begin{cases} \alpha(2s) & 0 \leq s \leq \frac{1}{2}, \\ \beta(2s-1) & \frac{1}{2} \leq s \leq 1. \end{cases}$$

The following figure illustrates this products of two paths.

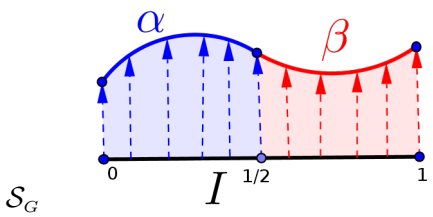


Figure 4.2: Figure illustrating the product of two paths.

An extension of the above definitions are also valid for loops. It is straightforward to see that the inverse of the product given above is defined by

$$(\alpha * \beta)^{-1}(s) = \beta^{-1} * \alpha^{-1}(s) = \begin{cases} \beta(1-2s) & 0 \leq s \leq \frac{1}{2}, \\ \alpha(2-2s) & \frac{1}{2} \leq s \leq 1. \end{cases}$$

Now we are ready to define homotopy. Given two loops α and β at the same base point x_0, we write $\alpha \sim \beta$ and say they are homotopic if there is a continuous map F such that

$$F : I \times I \to X$$
$$(s,t) \longmapsto F(s,t) = \begin{cases} F(s,0) = \alpha(s), \quad F(s,1) = \beta(s) & \forall s \in I \\ F(0,t) = F(1,t) = x_0 & \forall t \in I. \end{cases}$$

The map F is called a homotopy. We say that α is homotopic to β if $\alpha \sim \beta$. If α is not homotopic to β we write $\alpha \nsim \beta$. The following figures illustrate homotopic and non-homotopic loops.

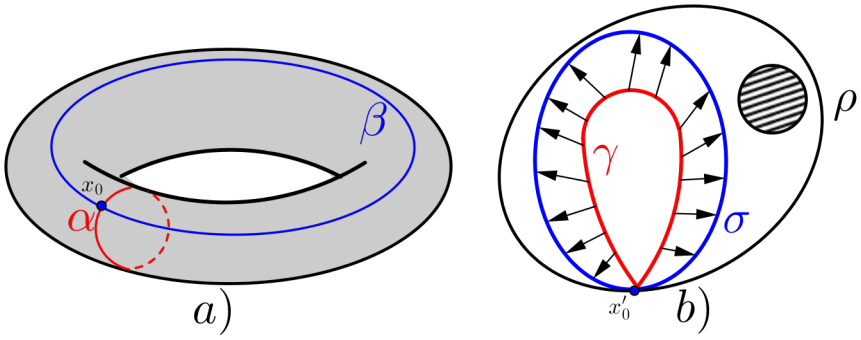


Figure 4.3: $a)$ We have $\alpha \nsim \beta$ on the torus T^2. The paths shown can not be deformed continuously into each other on the torus. $b)$ We have $\gamma \sim \sigma$ but $\gamma \nsim \rho$ and $\sigma \nsim \rho$ because of the hole in the 2-dimensional plane. Note that all these loops are homotopic to each other if they were embedded in 3-dimensional space and we can move past the hole by lifting the curves out of the page.

With the help of these figures, the notion of homotopic loops is clearly strongly related to the topology of the embedding space. It is also useful to define the connectedness of a topological space.

4.2.1 Definition. A topological space X is said to be *arcwise connected* or *connected* if we can define a path from I to X between any pair of points in X. Moreover X is said to be *simply connected* if any loop defined on X can be shrunk to a single point. As an example, the torus in figure $a)$ 4.3 is connected but not simply connected.

"Homotopic to" is an equivalence relation. Indeed, it is straightforward to show reflexivity $\alpha \sim \alpha$. The continuous map for this is defined by

$$\begin{cases} F(s,t) = \alpha(s) \ \forall t \in I \\ F(s,0) = F(s,1) = \alpha(0) \end{cases} .$$

Consider three loops $\alpha,\ \beta,\ \gamma$ on a topological space X. F is the homotopy map such that $\alpha \sim \beta$. The symmetric relation $\beta \sim \alpha$ follow from the homotopy map $F(s, 1-t)$. It is also straightforward to see the transitivity is defined by the homotopy map H such that

$$H(s,t) = \begin{cases} F(s,2t) & 0 \leq t \leq \frac{1}{2} \\ G(s,1-2t) & \frac{1}{2} \leq t \leq 1 \end{cases}$$

where $\beta \sim \gamma$ has homotopy $G(s,t)$.

On a topological space X, the set of loops homotopic to α is called the "homotopy class" of α and denoted by $[\alpha]$. Naturally the set of homotopy classes partition the set of loops at a point on X into disjoint subsets. The set of homotopy classes of loops at x_0 on X is denoted by $\pi_1(X, x_0)$ and is called the fundamental group of the topological space X. If the topological space X is connected then there is no need to specify the base point of the loops, so the fundamental group is simply written as $\pi_1(X)$. The group operation is defined as follow. Given a topological space X, two loops $\alpha,\ \beta$ on X, the group product is defined by

$$[\alpha] * [\beta] = [\alpha * \beta].$$

We must check that the right hand side does not depend on the representatives α and β. As an illustration, consider $F(s,t)$, $G(s,t)$ which are respectively the homotopy between $\alpha_1,\ \alpha_2 \in [\alpha]$ and

$\beta_1,\ \beta_2 \in [\beta]$. Thus, for a given t, the loop product $\alpha_1 * \beta_1$ can be defined by a continuous map $\tilde{H}_t(s)$ such that

$$\tilde{H}_t(s) = H(s,t) = \begin{cases} F(2s,t) & 0 \leq s \leq \frac{1}{2}, \forall t \in I \\ G(2s-1,t) & \frac{1}{2} \leq s \leq 1, \forall t \in I; \end{cases}$$

where $H(s,t)$ satisfies the conditions

$$H(0,t) = H(1,t) = (\alpha_1 * \beta_1)(0) = (\alpha_1 * \beta_1)(1) = x_0,\ \forall t \in I.$$

But by definition we have $\alpha_1 \sim \alpha_2$ and $\beta_1 \sim \beta_2$ thus $H(s,t)$ is promoted to the homotopy map ensuring $\alpha_1 * \beta_1 \sim \alpha_2 * \beta_2$. Hence, together $\alpha_1 * \beta_1 \sim \alpha * \beta$ and $\alpha_2 * \beta_2 \sim \alpha * \beta$ define a homotopy class of the loop product between $[\alpha]$ and $[\beta]$. This last statement shows that the class $[\alpha * \beta]$ of the product of a loop in $[\alpha]$ with a loop in $[\beta]$ is unique and doesn't depend on the representative of the classes that are multiplied. Thus, the class of the product of loops in X is the product of classes in $\pi_1(X, x_0)$. The unit element of the fundamental group $\pi_1(X, x_0)$ is $[c_{x_0}]$ the class of loops in X that can be shrunk to the base point. The inverse of an element $[\alpha]$ is given by $[\alpha]^{-1} = [\alpha^{-1}]$. Indeed, we have

$$\begin{aligned} [\alpha] * [c_{x_0}] = [c_{x_0}] * [\alpha] = [\alpha * c_{x_0}] = [\alpha], \\ [\alpha] * [\alpha]^{-1} = [\alpha]^{-1} * [\alpha] = [\alpha * \alpha^{-1}] = [c_{x_0}]. \end{aligned}$$

A generalization of a loop on a manifold X is called an n-loop on X at the base point x_0. Given a manifold X and an interval $I = [0,1]$, an n-loop on the manifold X is defined by a continuous map from $I^n = \underbrace{I \times I \cdots I}_{n}$ to X. Explicitly, an n-loop α at a base point $x_0 \in X$ is defined by

$$\begin{aligned} \alpha : I^n &\longrightarrow X \\ (t_1, \cdots, t_n) &\mapsto \alpha(t_1, \cdots, t_n), \end{aligned}$$

where

$$\alpha(t_i|_{i=1,\cdots,n} = 0) = \alpha(t_i|_{i=1,\cdots,n} = 1) = x_0.$$

By analogy with the idea of the first homotopy group, we can consider the n^{th} homotopy group of a manifold X, denoted by $\pi_n(X, x_0)$. Accordingly, $\pi_n(X, x_0)$ is a set of homotopy classes of n-loops at the base point x_0. We will focus on an n-dimensional sphere S^n as our n-loop. Note that there is no general algorithm to compute homotopy groups. Therefore, the computation of the homotopy groups of topological spaces must in general be carried out case by case.

4.3 Toy model problem

As mentioned at the beginning of the previous section, for our toy model problem we want to classify field configurations of finite action in 2-dimensional Euclidean space, where the gauge group is $G = U(1)$. As a starting point, we first want to define maps from S^1_∞ to G. Recall that an element of G must satisfy $gg^\star = |g|^2 = 1$. Thus, there must be a real θ modulo 2π such that $g = e^{i\theta}$. As a consequence G itself is isomorphic to S^1. Thus, the maps that we are looking for are endomorphisms of S^1 i.e. maps from S^1 to S^1. Clearly, our problem is now equivalent to finding the first homotopy group which is $\pi_1(S^1)$.

Consider the following maps

$$\begin{aligned} g^{(0)}(\theta) &= 1: \quad \text{the trivial map} \\ g^{(1)}(\theta) &= e^{i\theta}: \quad \text{the identity map} \\ g^{(\nu)}(\theta) &= e^{i\nu\theta}, \quad \text{where } \nu \in \mathbb{Z} \text{ called the winding number.} \end{aligned}$$

The trivial map is a constant map from S^1 to S^1, i.e. it maps S^1 to a single point of coordinate $\theta = 0$ on S^1. From its definition, $g^{(1)}$ maps each point of coordinate θ to itself. In contrast $g^{(\nu)}$ maps in such a way that varying θ from 0 to 2π we find that the image S^1 is traversed ν times.

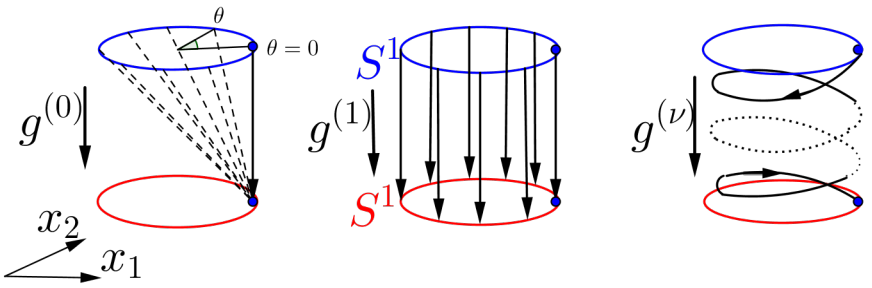


Figure 4.4: Figures illustrating the three possible maps from S^1 to S^1.

It is intuitively clear that any endomorphism of S^1 can be continuously deformed into one of the three maps above. For example, the fact that S^1 is connected implies that any constant map from S^1 to itself is homotopic to $g^{(0)}$. Consider now $g \in \pi_1(S^1)$. The winding number is given by

$$\nu = \frac{i}{2\pi}\int_0^{2\pi} d\theta g\left(\frac{dg^{-1}}{d\theta}\right) = \frac{i}{2\pi}\int_0^{2\pi} d\theta \frac{d\log(g^{-1})}{d\theta}. \tag{4.3.1}$$

To start off we will prove that (4.3.1) is invariant under continuous deformation of g (that is to say ν is only a function of homotopy class). To this end, consider the following deformation

$$g \to g + \delta g \Rightarrow g^\star \to g^\star + \delta g^\star.$$

The requirements that $gg^\star = 1$ and $(g+\delta g)(g^\star + \delta g^\star) = 1$ implies that there must be some number $\alpha \in \mathbb{R}$ modulo 2π such that $g + \delta g = e^{i\alpha}g$. An infinitesimal deformation of g forces us to take a small α. Now use (4.3.1) to compute the winding number of the deformed g.

We have

$$\begin{aligned} \nu|_{g+\delta g} &= \frac{i}{2\pi}\int_0^{2\pi} d\theta \frac{d\log(e^{-i\alpha}g^{-1})}{d\theta}, \\ &= \frac{i}{2\pi}\int_0^{2\pi} d\theta \left(\frac{d\log(g^{-1})}{d\theta} - \frac{d(i\alpha)}{d\theta}\right) = \frac{i}{2\pi}\int_0^{2\pi} d\theta \frac{d\log(g^{-1})}{d\theta}, \\ &= \nu|_g. \end{aligned}$$

This proves that the winding number is a homotopy invariant. Next we will prove that for two different winding numbers ν_1 and ν_2 the product map $g^{(\nu_1)}g^{(\nu_2)}$ has the winding number $\nu_1 + \nu_2$. We know that

$U(1)$ is an Abelian group. Thus, we have

$$\begin{aligned}
\nu|_{g^{(\nu_1)}g^{(\nu_2)}} &= \frac{i}{2\pi}\int_0^{2\pi} d\theta g^{(\nu_1)}g^{(\nu_2)}\frac{d}{d\theta}\left(g^{(\nu_1)}g^{(\nu_2)}\right)^{-1}, \\
&= \frac{i}{2\pi}\int_0^{2\pi} d\theta g^{(\nu_1)}g^{(\nu_2)}\frac{d}{d\theta}\left(g^{(\nu_2)^{-1}}g^{(\nu_1)^{-1}}\right), \\
&= \frac{i}{2\pi}\int_0^{2\pi} d\theta g^{(\nu_1)}\frac{dg^{(\nu_1)^{-1}}}{d\theta} + \frac{i}{2\pi}\int_0^{2\pi} d\theta g^{(\nu_2)}\frac{dg^{(\nu_2)^{-1}}}{d\theta}, \\
&= \nu_1 + \nu_2.
\end{aligned}$$

This is enough to prove that $g^{(\nu)}$ has winding number ν. To conclude our study of this toy model problem we will work out the relation between the winding number and the gauge field A_μ. Recall that we are in 2-dimensional Euclidean space. Thus, the index μ runs from 1 to 2. Consider a vector $G_\mu = \frac{i}{2\pi}\epsilon_{\mu\nu}A_\nu$ where $A_\mu = g\partial_\mu g^{-1}$ and $\epsilon_{12} = 1$. The radial unit vector in two dimensions is $\hat{r} = \cos(\theta)\hat{x}_1 + \sin(\theta)\hat{x}_2$. Thus, the inner product of the vectors G_μ and $\hat{r}_\mu$ is

$$\hat{r}_\mu G_\mu = \frac{i}{2\pi}\left(A_2\cos(\theta) - A_1\sin(\theta)\right) \tag{4.3.2}$$

Replacing A_μ with its definition in terms of $g \in U(1)$ yields

$$\hat{r}_\mu G_\mu = \frac{i}{2\pi}\left(\cos(\theta)g\frac{\partial g^{-1}}{\partial x_2} - \sin(\theta)g\frac{\partial g^{-1}}{\partial x_1}\right). \tag{4.3.3}$$

Next, it is also useful to recall the relations

$$\begin{cases} \frac{\partial x_1}{\partial\theta} &= -r\sin(\theta), \\ \frac{\partial x_2}{\partial\theta} &= r\cos(\theta). \end{cases}$$

Using these identities with the help of (4.3.3), we have

$$\hat{r}_\mu G_\mu = \frac{i}{2r\pi}\left(g\frac{\partial x_a}{\partial\theta}\frac{\partial g^{-1}}{\partial x_a}\right) = \frac{i}{2r\pi}g\frac{\partial g^{-1}}{\partial\theta}. \tag{4.3.4}$$

$$\Rightarrow \nu = \oint_{\partial D_r} d\theta r\hat{r}_\mu G_\mu = \int_{D_r} d^2x\partial_\mu G_\mu, \tag{4.3.5}$$

$$= \frac{i}{2\pi}\int_{D_r} d^2x\epsilon_{\mu\nu}\partial_\mu A_\nu, \tag{4.3.6}$$

$$= \frac{i}{4\pi}\int_{D_r} d^2x\epsilon_{\mu\nu}F_{\mu\nu}. \tag{4.3.7}$$

Above D_r is the 2-dimensional disk of radius r while ∂D_r is its boundary which is homotopic to a circle S^1. The divergence theorem was applied to obtain the last line of the above equation. Equation (4.3.7) shows how the winding number is directly related to the field strength in the two dimensional Euclidean space.

4.4 $G = SU(2)$ in 4-dimensional Euclidean Yang-Mills Theory

Consider now the case of gauge group $SU(2)$. This gauge group is isomorphic to the manifold S^3. Indeed, any element $g \in SU(2)$ can be written in terms of the Pauli matrices σ_i as

$$g = b_4\mathbb{1} + i\vec{b}\cdot\vec{\sigma}, \quad \text{where} \quad (b_4, b_1, b_2, b_3) \in \mathbb{R}^4. \tag{4.4.1}$$

The Pauli matrices satisfy the following identities

$$\sigma_i\sigma_j + \sigma_j\sigma_i = 2\delta_{ij}\mathbb{1}, \tag{4.4.2}$$

$$\sigma_i^\dagger = \sigma_i. \tag{4.4.3}$$

Thus, we have

$$gg^\dagger = (b_4 + i\vec{b}\cdot\vec{\sigma})(b_4 - i\vec{b}\cdot\vec{\sigma}) = \mathbb{1}, \tag{4.4.4}$$

$$= (b_4^2 + \vec{b}^2)\mathbb{1}, \tag{4.4.5}$$

$$\Rightarrow b_4^2 + b_1^2 + b_2^2 + b_3^2 = 1. \tag{4.4.6}$$

This last equation shows that there is indeed an isomorphism between $SU(2)$ and S^3. In other words, any point $(b_4, \vec{b})$ on S^3 defines an element g in $SU(2)$ and vice-versa. Now we have transformed our problem of classifying maps from S^3_∞ to $G = SU(2)$ into finding the homotopy group $\pi_3(S^3)$. Recall that the third homotopy group $\pi_3(S^3)$ of the three sphere S^3 is the set of classes of 3-loops on S^3. A simple example of a 3-loop is the sphere S^3 itself. By analogy with the construction of $\pi_1(S^1)$, it is not difficult see that the classification of the maps can be made by counting how many times they wrap the three sphere. The following figure shows a map from S^2 to S^2 with winding number 2.

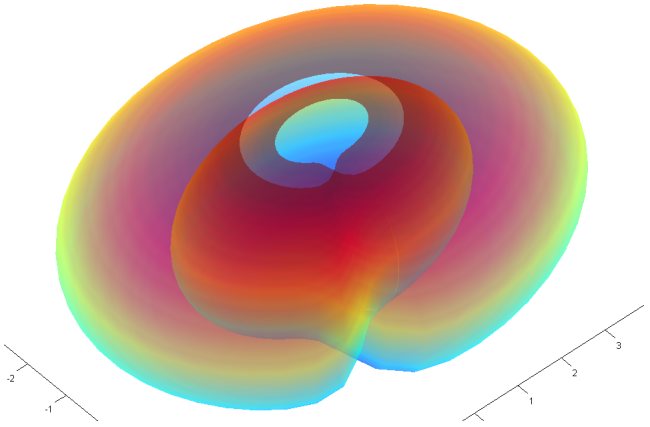


Figure 4.5: Profile section of a sphere S^2 wrapping itself twice.

Following our discussion of the elements of $\pi_1(S^1)$, we introduce the following standard mappings

$$g^{(0)}(x) = \mathbb{1} \quad \text{the trivial map,}$$
$$g^{(1)}(x) = (x_4\mathbb{1} + i\vec{x}\cdot\vec{\sigma}), \quad \text{where } |x|^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1$$
$$g^{(\nu)}(x) = [g^{(1)}(x)]^\nu, \quad \text{where } \nu \in \mathbb{Z}.$$

Note that x is always picked on the surface of S^3. Thus it always satisfies $|x|^2 = \sum_i x_i^2 = 1$. This time the case of ν negative is related to the orientation of the map wrapping the S^3. ν negative corresponds to the case that our map takes a left handed trio of tangent vectors into a right handed trio. Besides the name winding number, the integer ν is also called the Pontryagin class. The physical interpretation of $|\nu|$ is the same as the winding number found in the toy model problem. Indeed, regarding the maps defined above, it refers to the number of times the mapping wraps the three sphere S^3.

We will accept, without a proof, that every mapping $S^3 \to S^3$ is homotopic to $g^{(\nu)}$ for some integer ν. The formula for the winding number is given by

$$\nu = \frac{1}{24\pi^2}\oint d\theta_1 d\theta_2 d\theta_3 \epsilon^{ijk}\mathrm{tr}\left(g\frac{\partial g^{-1}}{\partial\theta_i}g\frac{\partial g^{-1}}{\partial\theta_j}g\frac{\partial g^{-1}}{\partial\theta_k}\right). \tag{4.4.7}$$

Let us prove that this winding number formula is independent of the coordinates θ_1, θ_2 and θ_3. To this end, we need to consider a change in coordinates. Consider the new coordinate $X_i(\theta_j)$, where $i, j = 1, 2, 3$. Thus, we have

$$\nu = \frac{1}{24\pi^2}\oint d\theta_1 d\theta_2 d\theta_3 \epsilon^{ijk}\mathrm{tr}\left(g\frac{\partial g^{-1}}{\partial\theta_i}g\frac{\partial g^{-1}}{\partial\theta_j}g\frac{\partial g^{-1}}{\partial\theta_k}\right), \tag{4.4.8}$$

$$= \frac{1}{24\pi^2}\oint d\theta_1 d\theta_2 d\theta_3 \epsilon^{ijk}\frac{\partial X_{\tilde{i}}}{\partial\theta_i}\frac{\partial X_{\tilde{j}}}{\partial\theta_j}\frac{\partial X_{\tilde{k}}}{\partial\theta_k}\mathrm{tr}\left(g\frac{\partial g^{-1}}{\partial X_{\tilde{i}}}g\frac{\partial g^{-1}}{\partial X_{\tilde{j}}}g\frac{\partial g^{-1}}{\partial X_{\tilde{k}}}\right). \tag{4.4.9}$$

Noting that

$$d\theta_1 d\theta_2 d\theta_3 = \left[\det\left(\frac{\partial X}{\partial\theta}\right)\right]^{-1} dX_1 dX_2 dX_3;$$

$$\epsilon^{ijk}\frac{\partial X_{\tilde{i}}}{\partial\theta_i}\frac{\partial X_{\tilde{j}}}{\partial\theta_j}\frac{\partial X_{\tilde{k}}}{\partial\theta_k} = \left[\det\left(\frac{\partial X}{\partial\theta}\right)\right]\epsilon^{\tilde{i}\tilde{j}\tilde{k}},$$

(4.4.9) becomes

$$\nu = \frac{1}{24\pi^2}\oint dX_1 dX_2 dX_3 \epsilon^{\tilde{i}\tilde{j}\tilde{k}}\mathrm{tr}\left(g\frac{\partial g^{-1}}{\partial X_{\tilde{i}}}g\frac{\partial g^{-1}}{\partial X_{\tilde{j}}}g\frac{\partial g^{-1}}{\partial X_{\tilde{k}}}\right). \tag{4.4.10}$$

Hence the winding number formula is indeed independent of the choice of coordinates. Next, we want to show that the winding number is also homotopy invariant. Consider a small change of an element $g = e^{\lambda^a T^a}$ of the group. We have

$$g(1 + \delta\lambda^a T^a) = g + g\delta T \Rightarrow \delta g = g\delta T. \tag{4.4.11}$$

The variation of g together with the identity $(g + \delta g)(g^{-1} + \delta g^{-1}) = \mathbb{1} \Rightarrow \delta g^{-1} = -g^{-1}\delta g g^{-1}$ yield

$$\delta(g\partial_k g^{-1}) = \delta g\partial_k g^{-1} + g\partial_k\delta g^{-1}, \tag{4.4.12}$$

$$= \delta g\partial_k g^{-1} - g\partial_k(g^{-1}\delta g g^{-1}), \tag{4.4.13}$$

$$= \cancel{\delta g\partial_k g^{-1}} - g\partial_k g^{-1}\delta g g^{-1} - \partial_k\delta g g^{-1} - \cancel{\delta g\partial_k g^{-1}}, \tag{4.4.14}$$

$$= -g\partial_k g^{-1}\delta g g^{-1} - \partial_k\delta g g^{-1}. \tag{4.4.15}$$

Plugging the relation $\partial_k(gg^{-1}) = 0 \Rightarrow \partial_k g g^{-1} = -g\partial_k g^{-1}$ into (4.4.15) and using (4.4.11) yields

$$\delta(g\partial_k g^{-1}) = \partial_k g\delta T g^{-1} - \partial_k(g\delta T)g^{-1}, \tag{4.4.16}$$

$$= \cancel{\partial_k g\delta T g^{-1}} - \cancel{\partial_k g\delta T g^{-1}} - g\partial_k\delta T g^{-1}, \tag{4.4.17}$$

$$= g\partial_k\delta T g^{-1}. \tag{4.4.18}$$

Now, insert this variation into the winding number formula and use the cyclicity of the trace and the identity $\partial_k g g^{-1} = -g\partial_k g^{-1}$ to find

$$\delta\nu = \frac{3}{24\pi^2}\oint d\theta_1 d\theta_2 d\theta_3 \epsilon^{ijk}\mathrm{tr}\big(g\partial_i g^{-1} g\partial_j g^{-1}\delta(g\partial_k g^{-1})\big), \tag{4.4.19}$$

$$= \frac{-1}{8\pi^2}\oint d\theta_1 d\theta_2 d\theta_3 \epsilon^{ijk}\mathrm{tr}\big(g\partial_i g^{-1} g\partial_j g^{-1} g\partial_k\delta T g^{-1}\big), \tag{4.4.20}$$

$$= \frac{-1}{8\pi^2}\oint d\theta_1 d\theta_2 d\theta_3 \epsilon^{ijk}\mathrm{tr}\big(\partial_i g^{-1}\partial_j g\partial_k\delta T\big), \tag{4.4.21}$$

$$= \frac{-1}{8\pi^2}\oint d\theta_1 d\theta_2 d\theta_3 \epsilon^{ijk}\partial_k\mathrm{tr}\big(\partial_i g^{-1}\partial_j g\delta T\big). \tag{4.4.22}$$

The passage to the last line has used the fact that ϵ^{ijk} is completely antisymmetric but $\partial_k\partial_i$ and $\partial_k\partial_j$ are symmetric. Thus, their contraction must vanish. Finally, using the divergence theorem, (4.4.22) is nothing but an integral on the boundary of the sphere S^3

$$\delta\nu = \frac{-1}{8\pi^2}\oint_{\partial S^3} d\hat{S}_k\epsilon^{ijk}\mathrm{tr}\big(\partial_i g^{-1}\partial_j g\delta T\big) = 0 \tag{4.4.23}$$

This integral is indeed zero because the three sphere S^3 doesn't have a boundary, i.e. $\partial S^3 = \emptyset$. Hence the variation of the winding number under a continuous deformation is zero. This statement can be rephrased as "the winding number is a function of homotopy classes".

Unlike the winding number formula for $\pi_1(S^1)$, which could be guessed, the winding number formula for $\pi_3(S^3)$ is more complicated. However, the winding number formula must involve integration of elements in the Lie algebra just as in $\pi_1(S^1)$ problem. Indeed, for any element g of the Lie group, $g\partial_\alpha g^{-1}$ is an element of the Lie algebra of the group. This is due to the property that any element of a Lie group is by definition the exponential of an element of the Lie algebra. In the following, instead of proving the winding number formula in (4.4.7), we will check by direct calculations that it indeed gives the expected result when evaluated on a standard mapping of S^3. To see this, it is straightforward to compute that the trivial map $g^{(0)}(x) = \mathbb{1}$ has zero winding number. Consider now the identity map $g^{(1)}(x) = x_4 + i\vec{x}\cdot\vec{\sigma}$, where we define $x_4^2 + x_1^2 + x_2^2 + x_3^2 = x_4^2 + \vec{x}^2 = 1$ to be coordinates of the unit three sphere S^3. It is clear that from its definition that $g^{(1)}$ maps S^3 uniformly onto itself. Thus, the integrand appearing in the formula for the winding number, using the formula (4.4.7), will not depend on the point on S^3 where we evaluate it. We define the north pole of S^3 to be the point with the coordinates

$$x_N = (x_4, x_1, x_2, x_3) = (1, 0, 0, 0).$$

Thus the variation of $g^{(1)}(x)$ at the north pole is only in the transverse space of x_4. Consequently, $\{\theta_i\}_{i=1,2,3}$ can be identified with the $\{x_i\}_{i=1,2,3}$, coordinates of S^3 in the neighborhood of the north pole. Thus, we have

$$\begin{aligned} g^{(1)}(x)\partial_i g^{(1)^{-1}}(x)\Big|_{x_N} &= (x_4 + i\vec{x}\cdot\vec{\sigma})\partial_i(x_4 - i\vec{x}\cdot\vec{\sigma})|_{x_N}\,, \\ &= -i\sigma^i. \end{aligned}$$

Now, insert these results into (4.4.7) to obtain

$$\begin{aligned} \nu|_{g^{(1)}} &= \frac{1}{24\pi^2}\oint d\theta_1 d\theta_2 d\theta_3\epsilon^{ijk}\mathrm{tr}[(-i)^3\sigma^i\sigma^j\sigma^k], \\ &= -i\frac{1}{24\pi^2}6\mathrm{tr}[\sigma^1\sigma^2\sigma^3]\oint d^3\Omega_3, \\ &= -i\frac{1}{4\pi^2}\mathrm{tr}\begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix}2\pi^2, \\ \Rightarrow \nu|_{g^{(1)}} &= 1. \end{aligned}$$

This is indeed what we expect to get for the winding number of the identity map since it wraps the sphere once. Now given two integers $\nu_1,\ \nu_2$ we want to evaluate the winding number of a general map $g = g^{(\nu_1)}g^{(\nu_2)}$. By a continuous deformation of the maps $g^{(\nu_1)},\ g^{(\nu_2)}$ we can define $\tilde{g}^{(\nu_1)}$ and $\tilde{g}^{(\nu_2)}$ such that

$$\tilde{g}^{(\nu_1)}(x) = \begin{cases} g^{(\nu_1)}(x) & \text{if } x_4 - \epsilon > 0, \\ \mathbb{1} & \text{if } x_4 + \epsilon < 0; \end{cases}$$

$$\tilde{g}^{(\nu_2)}(x) = \begin{cases} \mathbb{1} & \text{if } x_4 - \epsilon > 0, \\ g^{(\nu_2)}(x) & \text{if } x_4 + \epsilon < 0. \end{cases}$$

The hemisphere defined by $x_4 - \epsilon > 0$ can be chosen to be the northern hemisphere of S^3 and denoted by S^3_N. In contrast $x_4 + \epsilon < 0$ is chosen to be the southern hemisphere and denoted S^3_S. By definition the map $\tilde{g} = \tilde{g}^{(\nu_1)}\tilde{g}^{(\nu_2)}$ is homotopic to g and since we showed that the winding number is homotopy invariant, the winding number $\nu|_g$ must be equal to $\nu|_{\tilde{g}}$. Using the above argument, we have

$$\begin{aligned} \nu|_{g^{(\nu_1)}g^{(\nu_2)}} &= \frac{1}{24\pi^2}\oint_{S^3} d^3\Omega_3 \epsilon^{ijk}\mathrm{tr}\left[\tilde{g}^{(\nu_1)}\tilde{g}^{(\nu_2)})\partial_i(\tilde{g}^{(\nu_1)}\tilde{g}^{(\nu_2)})^{-1}(\tilde{g}^{(\nu_1)}\tilde{g}^{(\nu_2)}) \times \right. \\ &\qquad \left. \partial_j(\tilde{g}^{(\nu_1)}\tilde{g}^{(\nu_2)})^{-1}(\tilde{g}^{(\nu_1)}\tilde{g}^{(\nu_2)})\partial_k(\tilde{g}^{(\nu_1)}\tilde{g}^{(\nu_2)})^{-1}\right], \\ &= \frac{1}{24\pi^2}\int_{S^3_N} d^3\Omega_3 \epsilon^{ijk}\mathrm{tr}\big(g^{(\nu_1)}\partial_i {g^{(\nu_1)}}^{-1} g^{(\nu_1)}\partial_j {g^{(\nu_1)}}^{-1} g^{(\nu_1)}\partial_k {g^{(\nu_1)}}^{-1}\big) \\ &\quad + \frac{1}{24\pi^2}\int_{S^3_S} d^3\Omega_3 \epsilon^{ijk}\mathrm{tr}\big(g^{(\nu_2)}\partial_i {g^{(\nu_2)}}^{-1} g^{(\nu_2)}\partial_j {g^{(\nu_2)}}^{-1} g^{(\nu_2)}\partial_k {g^{(\nu_2)}}^{-1}\big), \\ \Rightarrow \nu|_{g^{(\nu_1)}g^{(\nu_2)}} &= \nu_1 + \nu_2. \end{aligned}$$

This is sufficient to prove that $g^{(\nu)}$ has winding number ν. Indeed, using this result iteratively, it is straightforward to see that the map $g^{(\nu)} = [g^{(1)}]^\nu$ has a winding number ν. Now there is no doubt that the winding number formula given in (4.4.7) is correct.

We are now interested in relating the winding number formula to the field strengths $\mathcal{F}_{\mu\nu}$, as we did in the toy model. To this end, define G_μ as

$$G_\mu = 4\epsilon_{\mu\nu\rho\sigma}\mathrm{tr}\left(\mathcal{A}_\nu\partial_\rho\mathcal{A}_\sigma + \frac{2}{3}\lambda\mathcal{A}_\nu\mathcal{A}_\rho\mathcal{A}_\sigma\right). \tag{4.4.24}$$

Recall that we are working in Euclidean space so that repeated Greek indices are summed from 1 to 4. In terms of the group element g, we note that G_μ can be expressed as

$$G_\mu = -\frac{4}{3\lambda^2}\epsilon_{\mu\nu\rho\sigma}\mathrm{tr}\left(g\partial_\nu g^{-1} g\partial_\rho g^{-1} g\partial_\sigma g^{-1}\right). \tag{4.4.25}$$

Indeed, substituting the gauge field $\mathcal{A}_\nu$ in (4.4.24) by $\frac{1}{\lambda}g\partial_\nu g^{-1}$, we find

$$\begin{aligned} G_\mu &= \frac{4}{\lambda^2}\epsilon_{\mu\nu\rho\sigma}\mathrm{tr}\left(g\partial_\nu g^{-1}\partial_\rho(g\partial_\sigma g^{-1}) + \frac{2}{3}g\partial g^{-1} g\partial_\rho g^{-1} g\partial_\sigma g^{-1}\right), \\ &= \frac{4}{\lambda^2}\epsilon_{\mu\nu\rho\sigma}\mathrm{tr}\left(g\partial_\nu g^{-1}\partial_\rho g\partial_\sigma g^{-1} + g\partial_\rho\partial_\sigma g^{-1} + \frac{2}{3}g\partial g^{-1} g\partial_\rho g^{-1} g\partial_\sigma g^{-1}\right), \\ &= -\frac{4}{3\lambda^2}\epsilon_{\mu\nu\rho\sigma}\mathrm{tr}\left(g\partial_\nu g^{-1} g\partial_\rho g^{-1} g\partial_\sigma g^{-1}\right). \end{aligned}$$

The passage to the last line has used the fact that $\epsilon_{\mu\nu\rho\sigma}\partial_\rho\partial_\sigma = 0$ and $\partial_\rho g = -g(\partial_\rho g^{-1})g$. It is also useful to show that

$$G_\mu = 2\epsilon_{\mu\nu\rho\sigma}\mathrm{tr}\left(\mathcal{A}_\nu\frac{1}{\lambda}\mathcal{F}_{\rho\sigma} - \frac{2}{3}\lambda\mathcal{A}_\nu\mathcal{A}_\rho\mathcal{A}_\sigma\right). \tag{4.4.26}$$

Starting from the right hand side and recalling the definition $\frac{1}{\lambda}\mathcal{F}_{\rho\sigma} = \partial_\rho \mathcal{A}_\sigma - \partial_\sigma \mathcal{A}_\rho + \lambda[\mathcal{A}_\rho, \mathcal{A}_\sigma]$ yields

$$\begin{aligned}
2\epsilon_{\mu\nu\rho\sigma}\text{tr}\left(\mathcal{A}_\nu \frac{1}{\lambda}\mathcal{F}_{\rho\sigma} - \frac{2}{3}\lambda \mathcal{A}_\nu \mathcal{A}_\rho \mathcal{A}_\sigma\right) &= 2\epsilon_{\mu\nu\rho\sigma}\text{tr}\left(\mathcal{A}_\nu \partial_\rho \mathcal{A}_\sigma - \mathcal{A}_\nu \partial_\sigma \mathcal{A}_\rho + \lambda \mathcal{A}_\nu [\mathcal{A}_\rho, \mathcal{A}_\sigma] - \lambda \frac{2}{3}\mathcal{A}_\nu \mathcal{A}_\rho \mathcal{A}_\sigma\right), \\
&= 4\epsilon_{\mu\nu\rho\sigma}\text{tr}\left(\mathcal{A}_\nu \partial_\rho \mathcal{A}_\sigma + \lambda \mathcal{A}_\nu \mathcal{A}_\rho \mathcal{A}_\sigma - \frac{\lambda}{3}\mathcal{A}_\nu \mathcal{A}_\rho \mathcal{A}_\sigma\right), \\
&= 4\epsilon_{\mu\nu\rho\sigma}\text{tr}\left(\mathcal{A}_\nu \partial_\rho \mathcal{A}_\sigma + \frac{2}{3}\lambda \mathcal{A}_\nu \mathcal{A}_\rho \mathcal{A}_\sigma\right), \\
&= G_\mu.
\end{aligned}$$

By analogy with (4.3.4) we introduce the unit vector $\hat{r}_\mu$, normal to the surface of S^3, so that

$$\nu = N \oint_{S^3} d^3 S \hat{r}_\mu G_\mu. \tag{4.4.27}$$

If we return to the expression for G_μ in (4.4.25), and note that the product $d^3 S \hat{r}_\mu G_\mu$ is nothing but an interior product between a vector and a 4-form, it becomes clear that the factor N in the previous expression must be equal to $-\frac{\lambda^2}{32\pi^2}$. Indeed, we have

$$\begin{aligned}
\nu = -\frac{\lambda^2}{32\pi^2}\oint_{S^3} d^3 S \hat{r}_\mu G_\mu &= -\frac{\lambda^2}{32\pi^2}\oint \left(-\frac{4}{3\lambda^2}\right) d^3 S \epsilon_{\nu\rho\sigma}\text{tr}\left(g\partial_\nu g^{-1} g\partial_\rho g^{-1}\partial_\sigma g^{-1}\right), \\
= \frac{1}{24\pi^2}\oint d^3 S \epsilon_{\nu\rho\sigma}\text{tr}\left(g\partial_\nu g^{-1} g\partial_\rho g^{-1}\partial_\sigma g^{-1}\right)
\end{aligned}$$

Thus, we have an alternative expression for the winding number

$$\nu = -\frac{\lambda^2}{32\pi^2}\oint_{S^3} d^3 S \hat{r}_\mu G_\mu. \tag{4.4.28}$$

This integral can be turned into a volume integral using Stoke's theorem and thus

$$\nu = -\frac{\lambda^2}{32\pi^2}\int d^4 x \partial_\mu G_\mu, \tag{4.4.29}$$

$$= -\frac{\lambda^2}{32\pi^2}\int d^4 x \epsilon_{\mu\nu\rho\sigma}\text{tr}\left(4\partial_\mu \mathcal{A}_\nu \partial_\rho \mathcal{A}_\sigma + 8\lambda \partial_\mu \mathcal{A}_\nu \mathcal{A}_\rho \mathcal{A}_\sigma\right). \tag{4.4.30}$$

The passage to the last line has used the expression for G_μ in (4.4.24). Consider the expansion of the product $\epsilon_{\mu\nu\rho\sigma}\text{tr}(\frac{1}{\lambda^2}\mathcal{F}_{\mu\nu}\mathcal{F}_{\rho\sigma})$ in terms of the gauge fields.

$$\begin{aligned}
\epsilon_{\mu\nu\rho\sigma}\text{tr}(\frac{1}{\lambda^2}\mathcal{F}_{\mu\nu}\mathcal{F}_{\rho\sigma}) &= \epsilon_{\mu\nu\rho\sigma}\text{tr}\big((\partial_\mu \mathcal{A}_\nu - \partial_\nu \mathcal{A}_\mu + \lambda[\mathcal{A}_\mu, \mathcal{A}_\nu])(\partial_\rho \mathcal{A}_\sigma - \partial_\sigma \mathcal{A}_\rho + \lambda[\mathcal{A}_\rho, \mathcal{A}_\sigma])\big), \\
&= \epsilon_{\mu\nu\rho\sigma}\text{tr}\big(4\partial_\mu \mathcal{A}_\nu \partial_\rho \mathcal{A}_\sigma + 8\lambda \partial_\mu \mathcal{A}_\nu \mathcal{A}_\rho \mathcal{A}_\sigma + \lambda^2 \mathcal{A}_\mu \mathcal{A}_\nu \mathcal{A}_\rho \mathcal{A}_\sigma\big).
\end{aligned}$$

The last term of this equation vanishes because of the cyclicity of the trace plus the antisymmetricity of the tensor $\epsilon_{\mu\nu\rho\sigma}$. Indeed, we have

$$\begin{aligned}
\epsilon_{\mu\nu\rho\sigma}\text{tr}(\mathcal{A}_\mu \mathcal{A}_\nu \mathcal{A}_\rho \mathcal{A}_\sigma) &= \epsilon_{\nu\rho\sigma\mu}\text{tr}(\mathcal{A}_\nu \mathcal{A}_\rho \mathcal{A}_\sigma \mathcal{A}_\mu), \\
&= \epsilon_{\nu\rho\sigma\mu}\text{tr}(\mathcal{A}_\mu \mathcal{A}_\nu \mathcal{A}_\rho \mathcal{A}_\sigma), \\
&= -\epsilon_{\nu\rho\mu\sigma}\text{tr}(\mathcal{A}_\mu \mathcal{A}_\nu \mathcal{A}_\rho \mathcal{A}_\sigma), \\
&= \epsilon_{\nu\mu\rho\sigma}\text{tr}(\mathcal{A}_\mu \mathcal{A}_\nu \mathcal{A}_\rho \mathcal{A}_\sigma), \\
&= -\epsilon_{\mu\nu\rho\sigma}\text{tr}(\mathcal{A}_\mu \mathcal{A}_\nu \mathcal{A}_\rho \mathcal{A}_\sigma), \\
&= 0.
\end{aligned}$$

The passage to the second line has used the cyclicity of the trace. We have also made use of the fact that $\epsilon_{\mu\nu\rho\sigma}$ is completely antisymmetric. Finally, the winding number formula, in terms of the field strengths is

$$\nu = -\frac{1}{32\pi^2}\int d^4x \epsilon_{\mu\nu\rho\sigma}\text{tr}(\mathcal{F}_{\mu\nu}\mathcal{F}_{\rho\sigma}). \tag{4.4.31}$$

Note that field configurations with different winding numbers can not be deformed into each other while keeping the action finite. Indeed, this follows because the winding number is invariant under any gauge transformation of the field configuration on the sphere S^3. Thus to change the winding number we must perform a transformation that is not a gauge transformation, implying that $\mathcal{F}_{\mu\nu}$ no longer vanishes as we go to infinity.

In the following discussion, we are interested in a general gauge group. We keep the space dimension fixed to $D = 4$. Recall that in (4.1.9), an element g of the gauge group G yields a field configuration on the surface of S^3_∞, of the form

$$\mathcal{A}_\mu = \frac{1}{\lambda} g\partial_\mu g^{-1}$$
$$\text{where} \quad g : S^3 \to G.$$

Thus, the homotopy class that we are interested in becomes $\pi_3(G)$. This is the homotopy group of mappings from S^3 into the Lie group G, which is by definition a manifold. It is useful to recall the following definitions:

- **A simple Lie group** is a connected non-Abelian Lie group such that its only normal subgroup is the trivial element $\{\mathbb{1}_G\}$.
- **A simple Lie algebra** is a non-Abelian Lie algebra whose only **ideals** are $\{\mathbb{0}_\mathfrak{g}\}$ and itself. A direct sum of simple Lie algebra is called a semi-simple Lie algebra.
- **A general compact Lie group** is the direct product of an Abelian group and simple groups. The following group product is an example of a general simple Lie group: $SU(2)^{\otimes q} \otimes SU(5)$, where $q \in \mathbb{Z}$.

The results that we have derived for the gauge group $G = SU(2)$ are valid for any simple Lie group G. This is due to Bott's proof that any continuous mapping : $S^3 \to G$ can be deformed into : $S^3 \to SU(2)$ where $SU(2)$ is a subgroup of G. In particular for $SU(n)$, it is straightforward to see that the entries of 2×2 blocks in $\mathfrak{su}(n)$ are the elements of $\mathfrak{su}(2)$. Thus, by embedding $SU(2)$ in $SU(n)$ the result seems quite natural.

Now, $\pi_3(U(1)) = \{\mathbb{1}_G\}$; this must be true because the only possible way of mapping a sphere S^3 onto a circle is by mapping all the points on S^3 onto one point on the circle. Hence the contribution of the Abelian factor of a compact Lie group always gives a winding number 0. However, due to Bott's result, each non-Abelian factor of a general simple Lie group must carry a non-trivial winding number. For example the homotopy group $\pi_3\big(U(1)^{\otimes p} \otimes SU(2)^{\otimes q} \otimes SU(5)\big)$ has winding number $(\nu_1, \ldots, \nu_q, \nu)$. Consequently, $\pi_3\big(U(1)^{\otimes p} \otimes SU(2)^{\otimes q} \otimes SU(5)\big)$ is isomorphic to $\mathbb{Z}^{q+1}$.

4.5 Evaluation of the path integral

We now want to carry out the semiclassical evaluation of

$$\int \mathcal{D}\mathcal{A}_\mu e^{-\frac{1}{4\lambda^2}\int d^4x \mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu}}.$$

Due to the gauge invariance of the theory, we have to choose a gauge fixing condition for the evaluation of the path integral. We will use axial gauge defined by $\mathcal{A}_3 = 0$. This gauge fixing condition is chosen because any non-singular field configuration can be put into axial gauge with a non-singular gauge transformation. As a consequence of gauge fixing, we need to include the Faddeev-Popov determinant in the path integral. Again we will evaluate this determinant by converting it into a path integral over ghosts fields. For an element of the gauge group $g \in G$, we recall that the gauge transformation of the field is

$$\mathcal{A}'_\mu = g\mathcal{A}_\mu g^{-1} + \frac{1}{\lambda} g\partial_\mu g^{-1}.$$

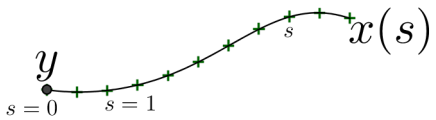
Thus, using the axial gauge fixing condition, we have

$$\begin{aligned}\frac{\delta F(\mathcal{A}'_\mu)}{\delta\chi} &= \frac{\delta g}{\delta\chi}\mathcal{A}_3 g^{-1} + \frac{1}{\lambda} g\mathcal{A}_3 \frac{\delta g^{-1}}{\delta\chi} + \frac{1}{\lambda}\frac{\delta}{\delta\chi}(g\partial_3 g^{-1}),\\ &= \frac{1}{\lambda}\frac{\delta}{\delta\chi}(g\partial_3 g^{-1}).\end{aligned}$$

This last line tells us that the Faddeev-Popov determinant is independent of the gauge fields $\mathcal{A}_\mu$. Thus, in this gauge the integral over the ghost field in the path integral can be absorbed into the measure $\mathcal{D}\mathcal{A}_\mu$.

Now, we want to show that any non-singular field configuration can indeed be put into axial gauge with a non-singular gauge transformation. Since the computation which follows is quite intricate, it is useful to summarize it in words. We will start by introducing a Wilson line. We show that the Wilson line satisfies a simple differential equation and using this differential equation we find the transformation law for a Wilson line under a gauge transformation. Armed with these results we write down a non-singular gauge transformation that moves us into the axial gauge. To this end, consider the Wilson-line defined by

$$U(x(s), y) = P\left[\exp\left(-\lambda\int_0^s ds' \frac{dx^\mu}{ds'} A^a_\mu[x(s')]T^a\right)\right],$$



where P is a path ordering notation defined as follows

$$P\big(\mathcal{O}(s)\tilde{\mathcal{O}}(s')\big) = \begin{cases}\mathcal{O}(s)\tilde{\mathcal{O}}(s') & s > s',\\ \tilde{\mathcal{O}}(s')\mathcal{O}(s) & s' > s.\end{cases}$$

This looks very similar to the time ordering notation of quantum mechanics. Accordingly, it is straightforward to see that the derivative with respect to s of the Wilson-line satisfies the differential equation

$$\frac{d}{ds}U(x(s),y) = -\lambda\frac{dx^\mu}{ds}A^a_\mu[x(s)]T^aU(x(s),y), \tag{4.5.1}$$

$$\Rightarrow \frac{dx^\mu}{ds}\big(\partial_\mu + \lambda\mathcal{A}_\mu[x(s)]\big)U(x(s),y) = 0. \tag{4.5.2}$$

Noting that the operator in brackets in the last line above is the covariant derivative, $D_\mu = \partial_\mu + \lambda\mathcal{A}_\mu$, in Euclidean space, we have

$$\frac{dx^\mu}{ds}D_\mu(\mathcal{A}_\mu)U(x(s),y) = 0. \tag{4.5.3}$$

The transformation law of the covariant derivative is

$$D'_\mu(\mathcal{A}'_\mu) = g(x)D_\mu(\mathcal{A}_\mu)g^{-1}(x).$$

The gauge transformation of the Wilson-line is given by

$$U'(x(s),y) = g(x)U(x(s),y)g^{-1}(y). \tag{4.5.4}$$

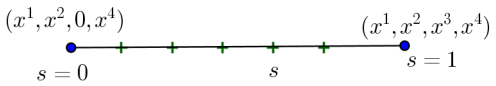
This can be proved by plugging the two last equations above in the differential equation that U satisfies. Indeed, we have

$$\begin{aligned}\frac{dx^\mu}{ds}D'_\mu(\mathcal{A}'_\mu)U'(x(s),y) &= \frac{dx^\mu}{ds}\left(g(x)D_\mu(\mathcal{A}_\mu)g^{-1}(x)\right)\left(g(x)U(x(s),y)g^{-1}(y)\right),\\ &= g(x)\left[\frac{dx^\mu}{ds}D_\mu(\mathcal{A}_\mu)U(x(s),y)\right]g^{-1}(y) = 0.\end{aligned}$$

Thus, $U'(x(s),y)$ satisfies the correct differential equation; moreover for $g = \mathbb{1}_G$ we have the identity $U'(x(s),y) = U(x(s),y)$. This prove that (4.5.4) is correct.

Now let P_x be the path generated by the tangent vector in the x^3 direction. Thus we have

$$\frac{dx^\mu}{ds} = \begin{cases} 0 & \mu \neq 3,\\ \neq 0 & \mu = 3.\end{cases} \tag{4.5.5}$$



A Wilson-line can be defined as

$$g(P_x) = P\exp\left(-\lambda\int_0^s ds'\frac{dx^\mu}{ds'}\mathcal{A}_\mu(x(s'))\right). \tag{4.5.6}$$

From the definition of a Lie group, we note that $g(P_x)$ is an element of the gauge group G. Now, perform a gauge transformation $g(P_x)^{-1}$ using (4.5.4) to obtain

$$\begin{aligned}g(P_x) \to g'(P_x) &= g^{-1}(P_x)g(P_x)g(P_{(x^1,x^2,0,x^4)}),\\ &= \mathbb{1}_G.\end{aligned}$$

The passage to the last line has used the fact that $g^{-1}(P_x)g(P_x) = \mathbb{1}_G$ and the definition of $g(P_x)$ evaluated at $s = 0$. In the new gauge

$$\frac{d}{ds}g'(P_x) = -\lambda\frac{dx^\mu}{ds}\mathcal{A}'_\mu(x(s))g'(P_x) = 0 \tag{4.5.7}$$

In other words, any non-singular field configuration can be indeed put into axial gauge with a non-singular gauge transformation.

In order to complete the evaluation of the path integral, we also need to fix the boundary conditions. We will work in a finite box of volume VT, where V is the three dimensional spacial volume and T is the Euclidean time. We will take $V, T \to \infty$ at the end. This will help us to understand how things depend on the boundary condition in the limit $V, T \to \infty$. The boundary conditions that we will use are those for which the surface terms arising in the derivation of the equations of motion vanish. The surface terms in the variation of the action are derived from

$$\begin{aligned}\delta S &\propto \int d^4x \mathrm{tr}(\partial_\mu \delta\mathcal{A}_\nu \mathcal{F}^{\mu\nu}),\\ &= \int d^4x \partial_\mu \mathrm{tr}(\delta\mathcal{A}_\nu\mathcal{F}^{\mu\nu}) - \int d^4x \mathrm{tr}(\delta\mathcal{A}_\nu\partial_\mu\mathcal{F}^{\mu\nu}).\end{aligned}$$

The passage to the last line has used an integration by parts. Using the divergence theorem, the total derivative term in the above equation is turned into a surface integral equal to

$$\int d^3S n_\mu \mathrm{tr}(\delta\mathcal{A}_\nu\mathcal{F}^{\mu\nu}).$$

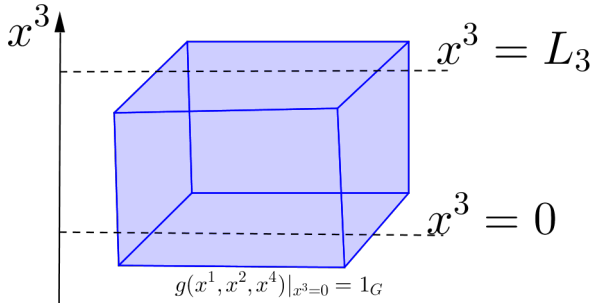
The normal components to the surface of $\mathcal{A}_\nu$ do not contribute because they are parallel to the vector n_μ, implying that $n_\mu\mathcal{A}_\nu$ is symmetric. The contraction of the antisymmetric $\mathcal{F}^{\mu\nu}$ with $n_\mu\delta\mathcal{A}_\nu$ must then vanish. Accordingly, the only relevant components of the field $\mathcal{A}_\mu$ are tangent to the surface. These are indeed the only components of $\mathcal{A}_\nu$ that we need in the evaluation of the winding number. The gauge fixing condition $\mathcal{A}_3 = 0$ must be satisfied by the tangential components and the boundary condition must yield a finite action as the box goes to infinity. For finite action, the field on the walls of the box is given by

$$\mathcal{A}_\mu = \frac{1}{\lambda}g\partial_\mu g^{-1}.$$

Using our gauge fixing condition, we must have

$$\frac{1}{\lambda}g\partial_3 g^{-1} = 0.$$

This implies that $\partial_3 g = 0$, in other words g is only a function of x^1, x^2 and x^4. Notice that we are still free to perform x^3 independent gauge transformations. To specify $\mathcal{A}_\mu$ on the boundary of the box we need to specify $g(x^1, x^2, x^4)$. Consider a gauge transformation which sets $g|_{x^3=0} = \mathbb{1}_G$.



Because of the gauge condition, $\partial_3 g = 0$ so that we have $g = \mathbb{1}_G$ on all sides of the box except on the side with $x^3 = L_3$. Thus, the full boundary condition is fixed by specifying g on the upper side of the box at $x^3 = L_3$. The following theorem will help us to understand the field configuration inside the box.

Theorem 4.5.1. *If boundary conditions g_1 and g_2 are in the same homotopy class, then any field configuration with boundary condition g_1 in an original box can be extended to a field configuration with boundary condition g_2 in a new box, consistent with the boundary conditions and $\mathcal{A}_3 = 0$ at the cost of an increase in action of $\frac{1}{\Delta}$. For a definition of what we mean by the old box and the new box, see Figure 4.6*

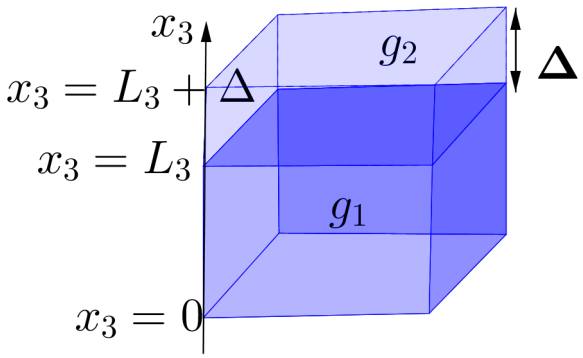


Figure 4.6: g_1 gives the boundary condition for the field configuration in the original box. g_2 gives the boundary condition for the field configuration in the bigger box. Δ is the size of the increase of the length of the side of the original box.

Proof. Let g_1 and g_2 be two mapping in the same homotopy class. The relevant homotopy obeys

$$g(x_1, x_2, s, x_4) = \begin{cases} g_1 & s = 0, \\ g_2 & s = 1. \end{cases}$$

Since g_1 and g_2 are respectively defined on $0 \leq x_3 \leq L_3$ and $L_3 \leq x_3 \leq L_3 + \Delta$, the argument s of the homotopy g can be defined to be $s = \frac{x_3 - L_3}{\Delta}$, thus

$$g(x_1, x_2, s, x_4) \equiv g(x_1, x_2, \frac{x_3 - L_3}{\Delta}, x_4).$$

It is clear that the field configuration $\mathcal{A}_\mu = \frac{1}{\lambda} g \partial_\mu g^{-1}$ extends the configuration from the original box into the new box. However, this extension does not preserve our gauge choice since $\mathcal{A}_3 \neq 0$. To overcome this define another field such that $\mathcal{A}_\mu = \begin{cases} \frac{1}{\lambda} g \partial_\mu g^{-1} & \mu \neq 3, \\ 0 & \mu = 3. \end{cases}$ This extension has $\mathcal{A}_3 = 0$ but is not pure gauge. Now, a perform a gauge transformation g^{-1}. We have

$$\mathcal{A}'_\mu = \begin{cases} g^{-1} \left(\frac{1}{\lambda} g \partial_\mu g^{-1} \right) g + \frac{1}{\lambda} g^{-1} \partial_\mu g = 0 & \mu \neq 3, \\ \frac{1}{\lambda} g^{-1} \partial_\mu g & \mu = 3. \end{cases}$$

The vanishing of the first case uses the identity $\partial_\mu g^{-1} g = -g^{-1} \partial_\mu g$. Since the dependence in x_3 of the mapping g is of the form $\frac{x_3 - L_3}{\Delta}$, we know that $\mathcal{A}'_3 = g^{-1} \partial_3 g(x_1, x_2, \frac{x_3 - L_3}{\Delta}, x_4)$ must be proportional to

$\frac{1}{\Delta}$. Accordingly, we have $\text{tr}(\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu}) \propto \frac{1}{\Delta^2}$. But the integral over the volume of the extra box produces a factor of Δ in the numerator. Thus, finally the total increase in the action has the form

$$S_\Delta = \frac{1}{4\lambda^2}\int_{\text{extra vol}} d^4x \text{tr}(\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu}) \propto \Delta\frac{1}{\Delta^2} = \frac{1}{\Delta}.$$

□

This proves that for a very large box, we can ignore the boundary conditions and simply integrate over all configuration with winding number $\nu = n$. For a given winding number, consider the transition amplitude between some initial configuration and final configuration

$$F(V,T,n) = N\int \mathcal{D}\mathcal{A}_1\mathcal{D}\mathcal{A}_2\mathcal{D}\mathcal{A}_4 e^{-S}\delta_{\nu n}. \tag{4.5.8}$$

Note that the initial and the final configuration are determined by the detailed boundary conditions that we impose on the fields. In general this transition amplitude must satisfy the following identity

$$F(V,T_1+T_2,n) = \sum_{n_1+n_2=n} F(V,T_1,n_1)F(V,T_2,n_2). \tag{4.5.9}$$

This formula is the discrete convolution of two transition amplitudes with different winding numbers. It is clear that (4.5.9) is not a vacuum to vacuum amplitude. In fact, our analysis of instantons in quantum mechanics in the first chapter showed that F is proportional to e^{-ET}, where E is the vacuum energy and T is the Euclidean time. Accordingly, the product of two transition amplitudes must be a product and not a convolution as in formula (4.5.9). The only way to resolve this issue is to work in the Fourier space dual to the winding number. Indeed, the Fourier image of the convolution of two functions is the product of the Fourier images of each function. Now, the discrete Fourier transformation is written as

$$F(V,T,\theta) = \sum_n e^{i2\pi n\theta}F(V,T,n). \tag{4.5.10}$$

(4.5.9) now becomes

$$F(V,T_1+T_2,\theta) = F(V,T_1,\theta)F(V,T_2,\theta). \tag{4.5.11}$$

This is now consistent with the expected structure $F \propto e^{-ET}$. One dramatic consequence of our argument is that the theory has many vaccua, labeled by θ. Now, using the definition of the transition amplitude and the above discussion, we have

$$\begin{aligned} F(V,T,\theta) &= \langle\theta|e^{-HT}|\theta\rangle, \\ &= N'\int \mathcal{D}\mathcal{A}_\mu e^{-S}e^{i2\pi\nu\theta}, \\ &= N'\int \mathcal{D}\mathcal{A}_\mu e^{-S+i2\pi\theta\nu}. \end{aligned}$$

From this last equation we see that the net effect of instantons is an additional term in the action

$$i2\pi\theta\nu = -\frac{i\theta}{16\pi}\int d^4x \epsilon_{\mu\nu\rho\sigma}\mathcal{F}_{\mu\nu}\mathcal{F}_{\rho\sigma}.$$

We will now describe a dramatic simplification in the task of constructing instanton and anti-instanton configurations. Let S_o be the action of an instanton. By definition, S_o is an extremum of the action. Define the dual field strength $\tilde{\mathcal{F}}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\mathcal{F}_{\rho\sigma}$. In terms of the dual field strength we write

$$S = \frac{1}{4\lambda^2}\int d^4x \mathrm{tr}(\mathcal{F}_{\mu\nu}\mathcal{F}_{\mu\nu}) = \frac{1}{2\lambda^2}\left(\int d^4x \mathrm{tr}(\mathcal{F}_{\mu\nu}\mathcal{F}_{\mu\nu})\int d^4x' \mathrm{tr}(\tilde{\mathcal{F}}_{\mu\nu}\tilde{\mathcal{F}}_{\mu\nu})\right)^{\frac{1}{2}}, \tag{4.5.12}$$

$$\geq \frac{1}{2\lambda^2}\left|\int d^4x \mathrm{tr}(\mathcal{F}_{\mu\nu}\tilde{\mathcal{F}}_{\mu\nu})\right| = \frac{1}{4\lambda^2}\left|\int d^4x \epsilon_{\mu\nu\rho\sigma}\mathrm{tr}(\mathcal{F}_{\mu\nu}\mathcal{F}_{\rho\sigma})\right|, \tag{4.5.13}$$

$$\Rightarrow S \geq \frac{8\pi^2}{\lambda^2}|\nu| \Rightarrow S_o = \frac{8\pi^2}{\lambda^2}. \tag{4.5.14}$$

It is straightforward to see that field configurations with fixed winding number are an extremum of the action if and only if the field strengths satisfy

$$\mathcal{F}_{\mu\nu} = \pm\frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\mathcal{F}_{\rho\sigma}. \tag{4.5.15}$$

Fields obeying this equation are known as dual or anti-selfdual field configurations respectively. Note that this is a first order partial differential equation. The equation of motions derived from Euler-Lagrange formalism are second order partial differential equations. We know that solution to (4.5.15) will solve the Euclidean equation of motion. The solutions with $\nu = 1$ have the smallest action. Using steepest-descent method, these are the only relevant solutions that we need in the semi-classical limit.

Now, we are interested in the evaluation of the transition amplitude for n instantons and $\tilde{n}$ anti-instantons by computing the path integral using the steepest descent method. In order to do this, we will again assume a dilute instanton gas. In other words, we assume that there is no overlap between field configurations of different winding numbers. Since we have already carried out this calculation in quantum mechanics in chapter two, we simply give the result. We have

$$\langle\theta|e^{-HT}|\theta\rangle \propto \int_0^{2\pi} d\theta \sum_{n,\tilde{n}} (Ke^{-S_o})^{n+\tilde{n}}\frac{(VT)^{n+\tilde{n}}}{n!\tilde{n}!}e^{i2\pi(n-\tilde{n})\theta}, \tag{4.5.16}$$

$$= \int_0^{2\pi} d\theta \exp\left(2KVTe^{-S_o}\cos(2\pi\theta)\right) \tag{4.5.17}$$

Hence, after identifying the argument of the exponential the vacuum energy density is given by

$$\frac{E(\theta)}{V} = -2K\cos(2\pi\theta)e^{-S_o}. \tag{4.5.18}$$

By translation invariance the vacuum expectation value (VEV) satisfies

$$\begin{aligned}
\langle\theta|\epsilon_{\mu\nu\rho\sigma}\mathcal{F}_{\mu\nu}\mathcal{F}_{\rho\sigma}|\theta\rangle &= \frac{1}{VT}\int d^4x \langle\theta|\epsilon_{\mu\nu\rho\sigma}\mathcal{F}_{\mu\nu}\mathcal{F}_{\rho\sigma}|\theta\rangle, \\
&= -\frac{32\pi^2}{VT}\frac{\nu\int\mathcal{D}\mathcal{A}_\mu e^{-S+i2\pi\nu\theta}}{\int\mathcal{D}\mathcal{A}_\mu e^{-S+i2\pi\nu\theta}}, \\
&= \frac{i16\pi}{VT}\frac{\partial}{\partial\theta}\log\left(\int\mathcal{D}\mathcal{A}_\mu e^{-S+i2\pi\nu\theta}\right), \\
&= \frac{i16\pi}{VT}\frac{\partial}{\partial\theta}\left(-2KVTe^{-S_o}\cos(2\pi\theta)\right), \\
&= -i64\pi^2 Ke^{-S_o}\sin(2\pi\theta).
\end{aligned}$$

The passage to the last line has used the result found in (4.5.17). The θ dependence in the VEV is firm evidence that θ labels distinct vaccua. Note also that our result is independent of V, T as expected. Finally, the VEV must be imaginary in Euclidean space so that after Wick rotation $x_4 \to ix_0$ the VEV is real in Minkowski space. Indeed, we have

$$\begin{aligned} x_4 \to ix_0 \Rightarrow \mathcal{F}_{j4} &= i\mathcal{F}_{j0}. \\ \Rightarrow \epsilon_{\mu\nu\rho\sigma}\mathcal{F}_{\mu\nu}\mathcal{F}_{\rho\sigma} &= \sum_{\sigma\in S_4} \mathsf{sign}(\sigma)\mathcal{F}_{\sigma(1)\sigma(2)}\mathcal{F}_{\sigma(3)\sigma(4)}, \\ &= i\sum_{\sigma\in S_4} \mathsf{sign}(\sigma)\mathcal{F}_{\sigma(1)\sigma(2)}\mathcal{F}_{\sigma(3)\sigma(0)}. \end{aligned}$$

Using the expression for the vacuum energy given in (4.5.18), it is straightforward to see that the $\nu > 1$ solutions are not needed for our semiclassical calculation. Indeed, the dilute instanton gas approximation implies that you know the $\nu > 1$ solutions as well as needed. To see this, recall that the action of a single instanton is

$$S_o = \frac{8\pi^2}{\lambda^2}. \tag{4.5.19}$$

The bi-instanton action is at least double the value of S_o and thus

$$S_o' = 2\frac{8\pi^2}{\lambda^2}. \tag{4.5.20}$$

It follows that the vacuum energy density becomes

$$\frac{E(\theta)}{V} = -2K\cos(2\pi\theta)e^{-S_o} - 2K'\cos(4\pi\theta)e^{-S_o'}.$$

We can already see that the bi-instaton term is exponentially damped. This implies that the effects of all the higher instanton number ($\nu \geq 2$) configurations are negligible compared to the contribution of the single instanton configurations.

4.6 Instanton solutions for Euclidean $SU(2)$ Yang-Mills theory

In this section we will set the coupling constant to 1 in order to simplify our discussion. Start with the basic solution for a single instanton i.e. winding number $\nu = 1$. The solution for a single instanton for the gauge group $SU(2)$ in 4 dimensional Euclidean space has been found by Belavin, Polyakov, Schwartz and Tyupkin [17]. The solution has the form

$$\mathcal{A}_\mu = \frac{x^2}{x^2+1} g^{(1)}\partial_\mu[g^{(1)}]^{-1}, \qquad g^{(1)} = \frac{x_4 + i\vec{x}\cdot\vec{\sigma}}{x}, \quad x^2 = x_4^2 + \vec{x}^2. \tag{4.6.1}$$

Note that this solution is consistent with the behaviour of the potential field as a pure gauge field on the three sphere at infinity

$$\mathcal{A}_\mu \underset{r\to\infty}{\longrightarrow} g^{(1)}\partial_\mu[g^{(1)}]^{-1}.$$

The solution given in (4.6.1) gives an instanton configuration with size normalized to 1. It is straightforward to see that this solution is not invariant under translation. The translation transformation will give

us another solution with the center of the instanton translated. In addition, (4.6.1) is not conformaly invariant. Indeed, the scaling transformation $x_\mu \longrightarrow \frac{x_\mu}{\rho}$ yields an inequivalent solution

$$\mathcal{A}_\mu = \frac{x^2}{x^2+\rho^2} g^{(1)} \partial_\mu [g^{(1)}]^{-1}. \tag{4.6.2}$$

The size of the instanton has been rescaled to ρ. We have thus identified five parameters which label different solutions. These parameters describe the center and the size of the instanton. We will now argue that our solution is invariant under rotations. The x^2 dependence is unchanged under a rotation. Under a rotation $g^{(1)}$ transforms as

$$g^{(1)} \longrightarrow g g^{(1)} h^{-1}, \quad g, h \in SU(2) \text{ are constant matrices.}$$

To understand this formula, recall that $SO(4) \simeq SU(2) \otimes SU(2)$. In summary, under a rotation our field transforms as

$$\begin{aligned} \mathcal{A}'_\mu &= \frac{x^2}{x^2+\rho^2} g g^{(1)} h^{-1} \partial_\mu \big(h [g^{(1)}]^{-1} g^{-1}\big), \\ &= \frac{x^2}{x^2+\rho^2} g \big(g^{(1)} \partial_\mu [g^{(1)}]^{-1}\big) g^{-1}. \end{aligned}$$

Hence, a rotation is equivalent to a constant gauge transformation. Since the $SU(2)$ gauge group has 3 parameters, constant gauge transformations generate 3 additional parameters which label different solutions. The next transformation that we will discuss is inversion. Our discussion follows [18, 19]. The coordinate transformation implementing an inversion is given by

$$x_\mu \longrightarrow \tilde{x}_\mu = \frac{x_\mu}{x^2}.$$

The transformation matrix $I_{\mu\nu}$ associated with this transformation is

$$\frac{\partial}{\partial x_\mu} = \frac{\partial \tilde{x}_\nu}{\partial x_\mu} \frac{\partial}{\partial \tilde{x}_\nu} \Leftrightarrow \partial_\mu = I_{\mu\tilde{\nu}} \partial_{\tilde{\nu}}.$$

Accordingly, we have

$$\begin{aligned} I_{\mu\nu} &= \frac{\partial}{\partial x_\mu} \left(\frac{x_\nu}{x_\rho x_\rho} \right), \\ &= \frac{\delta_{\mu\nu}}{x^2} - 2 \frac{x_\nu \delta_{\mu\rho} x_\rho}{x^4}, \\ &= \frac{1}{x^2} \left(\delta_{\mu\nu} - 2 \frac{x_\mu x_\nu}{x^2} \right). \end{aligned}$$

Thus, under inversion

$$\mathcal{A}_\mu \longrightarrow \mathcal{A}'_\mu = I_{\mu\nu} \mathcal{A}_\nu, \tag{4.6.3}$$

$$= \frac{1}{x^2} \left(\delta_{\mu\nu} - 2 \frac{x_\mu x_\nu}{x^2} \right) \frac{x^2}{1+x^2} g^{(1)} \partial_\nu [g^{(1)}]^{-1}, \tag{4.6.4}$$

$$= \frac{1}{1+x^2} \left(g^{(1)} \partial_\mu [g^{(1)}]^{-1} - 2 \frac{x_\mu x_\nu}{x^2} g^{(1)} \partial_\nu [g^{(1)}]^{-1} \right), \tag{4.6.5}$$

$$= \frac{1}{x^2+1} g^{(1)} \partial_\nu [g^{(1)}]^{-1}. \tag{4.6.6}$$

The passage to the last line has used the fact that $x_\nu \partial_\nu g^{(1)} = 0$. To see this, note that by dimensional analysis we expect $x_\nu \partial_\nu \propto r\frac{\partial}{\partial r}$, with $r^2 = x_4^2 + \vec{x}\cdot\vec{x}$. Indeed, the following example in two dimensions is in agreement with this simple argument.

$$\begin{aligned} & x = r\cos(\theta) \qquad y = r\sin(\theta), \\ \Rightarrow \frac{\partial}{\partial r} &= \frac{\partial x}{\partial r}\frac{\partial}{\partial x} + \frac{\partial y}{\partial r}\frac{\partial}{\partial y}, \\ &= \cos(\theta)\frac{\partial}{\partial x} + \sin(\theta)\frac{\partial}{\partial y}, \\ &= \frac{1}{r}\vec{x}\cdot\vec{\nabla} \quad \Rightarrow \quad \vec{x}\cdot\vec{\nabla} = r\frac{\partial}{\partial r}. \end{aligned}$$

We can also say that $x_\nu\partial_\nu$ is related to the directional derivative in the direction of x_ν which points radially. Hence $x_\nu\partial_\nu g^{(1)} = 0 = x_\nu\partial_\nu[g^{(1)}]^{-1}$. The gauge field found in (4.6.6) is the solution for an anti-instanton of winding number $\nu = -1$. It turns out that the behaviour of the anti-instanton on the surface of the three sphere at infinity has been lost. However, if we perform a gauge transformation of (4.6.6) by $U = [g^{(1)}]^{-1}$ we will correct this feature. Performing the transformation we find

$$\begin{aligned} \mathcal{A}'_\mu \longrightarrow & \frac{1}{1+x^2} U g^{(1)} \partial_\mu [g^{(1)}]^{-1} U^{-1} + U\partial_\mu U^{-1}, \\ &= \frac{1}{1+x^2}\partial_\mu [g^{(1)}]^{-1} g^{(1)} + [g^{(1)}]^{-1}\partial_\mu g^{(1)}, \\ &= \frac{x^2}{1+x^2}[g^{(1)}]^{-1}\partial_\mu g^{(1)}. \end{aligned}$$

This proves that there is no new parameter generated by inversion. Moreover, since the special conformal transformation is a combination of translations and inversions, we will not find new parameters from the special conformal transformation of our solution. We can therefore conclude that the total number of parameters for a single (anti- or) instanton is 8. There are exactly 5 parameters for the center and size of the instanton and 3 parameters associated to constant gauge transformations. Consequently there will be 8 zero modes, one associated to each parameter.

We will now demonstrate in detail that single instanton and anti-instanton solutions are invariant under rotations. A small generalization would extend the argument to special conformal transformations. To start off, recall the 't Hooft ansatz for the field strength

$$\mathcal{F}_{\mu\nu} = \frac{4i}{(1+x^2)^2}\sigma_{\mu\nu},$$

where $\sigma_{\mu\nu}$ are defined to be antisymmetric and satisfy

$$\begin{aligned} \sigma_{ij} &= -\frac{i}{4}[\sigma_i, \sigma_j], \qquad \sigma_i \text{ are the Pauli matrices;} \\ \sigma_{i4} &= \frac{1}{2}\sigma_i. \end{aligned}$$

These field strengths are selfdual. Indeed, quick calculation shows that $\sigma_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta}\sigma_{\alpha\beta}$. By definition, we have

$$\begin{aligned} \sigma_{12} \equiv -\frac{i}{4}[\sigma_1, \sigma_2] &= \frac{-i}{4}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} - \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right], \\ &= \frac{-i}{2}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} = \frac{1}{2}\sigma_3. \end{aligned}$$

Now, we compute

$$\frac{1}{2}\epsilon_{12\alpha\beta}\sigma_{\alpha\beta} = \frac{1}{2}(\frac{1}{2}\epsilon_{1234}\sigma_{34} + \frac{1}{2}\epsilon_{1243}\sigma_{43}) = \frac{1}{2}\sigma_3.$$

Proceeding in this manner we could verify that $\sigma_{\mu\nu}$ is indeed selfdual. Starting from the selfdual solution, it is straightforward to see that the anti-selfdual solutions are given by

$$\begin{aligned}\mathcal{F}_{\mu\nu} &= \frac{4i}{1+x^2}\bar{\sigma}_{\mu\nu},\\ \bar{\sigma}_{ij} &= \sigma_{ij},\\ \bar{\sigma}_{i4} &= -\frac{1}{2}\sigma_i.\end{aligned}$$

Recall that the above solutions for selfdual and anti-selfdual field strengths correspond, respectively, to a single instanton and anti-instanton. For simplicity we consider gauge group $O(4) \simeq SU(2) \times SU(2)$ which allows us to pair the instanton with the anti-instanton into a single configuration. To this end, we define

$$\mathcal{F}_{\mu\nu} = \frac{4i}{(1+x^2)^2}\Sigma_{\mu\nu}, \tag{4.6.7}$$

where $\Sigma_{\mu\nu}$ is defined in terms of Dirac's matrices as

$$\begin{aligned}\Sigma_{\mu\nu} &= \frac{-i}{4}[\alpha_i, \alpha_j] = \begin{pmatrix} \sigma_{\mu\nu} & 0 \\ 0 & \bar{\sigma}^{\mu\nu} \end{pmatrix},\\ \alpha_i &= \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix}, \quad \alpha_4 = i\begin{pmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}.\end{aligned}$$

Next, it is easy to see that (4.6.7) can be derived from the gauge field

$$\mathcal{A}_\mu = \frac{-2i}{1+x^2}\Sigma_{\mu\nu}x_\nu. \tag{4.6.8}$$

Under an infinitesimal rotation of parameter $\omega^{\alpha\beta}$, the change in the coordinates is $x_\mu \longrightarrow x_\mu - \omega_{\mu\alpha}x_\alpha$. Consequently, the net effect of the rotation on the field strength is given by

$$\delta\mathcal{F}_{\mu\nu} = \frac{4i}{(1+x^2)^2}\big(\omega_{\mu\alpha}\Sigma_{\alpha\nu} - \omega_{\nu\rho}\Sigma_{\rho\mu}\big). \tag{4.6.9}$$

However, an $O(4)$ gauge transformation $U = e^{-i\Theta}$, $\Theta = \frac{1}{2}\omega_{\mu\nu}\sigma_{\mu\nu}$ transforms $\mathcal{F}_{\mu\nu}$ into $U\mathcal{F}_{\mu\nu}U^{-1}$ and thus the variation at first order in Θ of the field strengths is

$$\delta\mathcal{F}_{\mu\nu} = i[\Theta, \mathcal{F}_{\mu\nu}], \tag{4.6.10}$$

$$= -\frac{4i}{(1+x^2)^2}[\Theta, \Sigma_{\mu\nu}]. \tag{4.6.11}$$

For the equation (4.6.11), we must again pair the selfdual (instanton) and anti-selfdual (anti-instanton) field strengths so that the gauge parameter Θ will be, after including $\bar{\sigma}_{\mu\nu}$, proportional to $\Sigma_{\mu\nu}$. After doing so, it is trivial to show that (4.6.9) and (4.6.11) can cancel each other. In conclusion, for any rotation of parameter $\omega^{\mu\nu}$ there is some gauge transformation $U \in O(4)$ such that the total net effect of these two transformations in the field strength is zero. This is why rotations do not produce zero modes of the instanton and anti-instanton. After a mild generalization one can show that for the special conformal transformation, i.e. inversion plus translation, the invariance of the field configuration is also accomplished by a gauge transformation.

4.7 Degrees of freedom for n instantons

In this section we determine how many parameter label solutions describing n instantons. Naively, since single instanton solutions are labeled by 8 parameters, we expect that there should be $8n$ parameters labeling solutions for n instantons. Indeed, we will show that in general, there are $8n$ parameters labeling the n instanton solutions. Towards this end, we recall the Witten and 't Hooft solutions for any arbitrary instanton number n for the $SU(2)$ gauge group. The gauge field ansatz for a selfdual solution is

$$\mathcal{A}_\mu = i\sigma_{\mu\nu}a^\nu, \tag{4.7.1}$$

where the vector field a^μ can be derived from a scalar potential ρ,

$$a_\mu = \partial_\mu \ln \rho. \tag{4.7.2}$$

Accordingly, the selfdual equation $\mathcal{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta}\mathcal{F}^{\alpha\beta}$ is reduced to

$$\frac{1}{\rho}\partial_\mu\partial^\mu\rho = 0. \tag{4.7.3}$$

For a given winding number n, a general solution of (4.7.3) has the form

$$\rho = \sum_{i=1}^{n+1} \frac{\lambda_i^2}{|x-y_i|^2}. \tag{4.7.4}$$

The right hand side of this equation seems to contain $5(n+1)$ parameters for the positions and sizes of the n instantons. However a common rescaling of the size of instantons leaves a_μ unchanged reducing the number of parameters to $5n+4$.

Now, consider a small variation for the gauge field

$$\mathcal{A}_\mu \to \mathcal{A}_\mu + \delta\mathcal{A}_\mu.$$

Based on the expression for the gauge field given in (4.7.1), the most general form for this variation is

$$\delta\mathcal{A}_\mu = i\sigma_{\alpha\beta}X_\mu{}^{\alpha\beta}, \tag{4.7.5}$$

where $X_\mu{}^{\alpha\beta}$ must be antisymmetric and selfdual in the indices $\alpha\beta$. This requirement reflects the fact that $\sigma_{\alpha\beta}$ is selfdual and $\mathcal{A}_\mu$ is antihermitian. Since $X_{\mu\alpha\beta}$ is antisymmetric in $\alpha\beta$, for each value of the index μ, there are $\frac{4(4-1)}{2} = 6$ independent components for each μ. The selfdual condition reduces this number by a half. Thus, in total $X_{\mu\alpha\beta}$ has twelve independent parameters. In addition, because of the gauge invariance of the small variation of $\mathcal{A}_\mu$, we need to subtract the number of the generators of the gauge group $SU(2)$. Finally, there are nine independent components of the matrix $X_{\mu\alpha\beta}$. The explicit form of the small variation of $\mathcal{A}_\mu$ in terms of the superpotential is

$$\delta\mathcal{A}_\mu = i\sigma_{\mu\nu}\partial^\nu\frac{\delta\rho}{\rho}. \tag{4.7.6}$$

Putting (4.7.5) and (4.7.6) together, we can set

$$\delta\mathcal{A}_\mu = i\sigma_{\mu\nu}\partial^\nu\frac{\delta\rho}{\rho} + i\rho\sigma_{\alpha\beta}\partial_\nu Y_\mu{}^{\nu\alpha\beta}, \tag{4.7.7}$$

where $Y_{\mu\nu\alpha\beta}$ must inherit the properties of $X_{\mu\alpha\beta}$, must satisfy the selfdual condition in the $\mu\nu$ indices and is subject to the condition

$$Y_{\mu\nu}{}^{\mu\nu} = 0. \tag{4.7.8}$$

Together, (4.7.7) and (4.7.8) imply that the number of independent functions is nine. Next, decompose $Y_{\mu\nu\alpha\beta}$ into symmetric, traceless and antisymmetric part as follows.

$$Y_{\mu\nu\alpha\beta} = S_{\mu\nu\alpha\beta} + A_{\mu\nu\alpha\beta},$$

where

$$\begin{aligned} S_{\mu\nu\alpha\beta} &= S_{\alpha\beta\mu\nu}, \\ A_{\mu\nu\alpha\beta} &= -A_{\alpha\beta\mu\nu}, \end{aligned}$$

It is straightforward to see that we can write

$$A_{\mu\nu\alpha\beta} = \frac{1}{4}(\delta_{\mu\alpha}V_{\nu\beta} - \delta_{\nu\alpha}V_{\mu\beta} - \delta_{\mu\beta}V_{\nu\alpha} + g_{\nu\beta}V_{\mu\alpha}),$$

where $V_{\alpha\beta}$ is antisymmetric and selfdual. Following the arguments of Jackiw and Rebbi [20], one finds that a small variation of the selfdual equation $\delta\mathcal{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta}\delta\mathcal{F}_{\alpha\beta}$ implies that the symmetric part of $Y_{\mu\nu\alpha\beta}$ satisfies

$$\partial_\rho\partial^\rho S_{\mu\nu\alpha\beta} = 0. \tag{4.7.9}$$

All non-trivial solutions of this equation have singularities which cannot be removed by gauge transformations. Indeed, there is no symmetric tensor among the possible parameters of the gauge group. Thus, the only physically meaningful solution of (4.7.9) is $S_{\mu\nu\alpha\beta} = 0$.

When the symmetric part of $Y_{\mu\nu\alpha\beta}$ vanishes, it has been shown in detail in [20] that singularities that arise from $V_{\mu\nu}$ can be removed by gauge transformations. The self dual condition written in terms of $\delta\rho$ and $V_{\mu\nu}$ are

$$\begin{aligned} \partial_\lambda\partial^\lambda\rho V_{\mu\nu} &= 0, \\ \partial_\lambda\partial^\lambda\delta\rho + 2\rho\partial_\mu V^\mu{}_\nu\partial^\nu\rho &= 0. \end{aligned}$$

The solutions of these equations for n instantons are

$$\begin{aligned} V_{\mu\nu} &= \frac{1}{\rho}\sum_{i=1}^{n+1}\frac{k^i_{\mu\nu}}{(x-y^i)^2}, \\ \delta\rho(x) &= \int d^4x' \frac{1}{4\pi^2(x-x')^2}(2\rho\partial_\mu V^\mu{}_\nu\partial^\nu\rho)(x') + \delta\tilde{\rho}(x), \end{aligned}$$

where the $k^i_{\mu\nu}$'s are constant antisymmetric and selfdual matrices and $\delta\tilde{\rho}$ is the harmonic variation of $\rho(x)$. These solutions have $8(n+1) - 1 - 10 = 8n - 3$ physical parameters. Adding back 3 parameters for global gauge transformations we finally obtain $8n$ parameters as we expected.

4.8 Integration over the scaling zero mode

In this section we will focus on the single instanton solution and we will discuss the integration over the zero mode corresponding to the size ρ of the instanton. To begin with, return to the number K in (4.5.18) given by the square root of the product of eigenvalues of a certain partial differential equation. We note by dimensional analysis that

$$\frac{E(\theta)}{V} = A\cos(2\pi\theta)e^{-\frac{8\pi^2}{\lambda^2}}\frac{1}{\lambda^8}\int_0^\infty \frac{d\rho}{\rho^5} f(\rho\mu). \tag{4.8.1}$$

The factor A is a constant of proportionality while $\cos(2\pi\theta)$ reflects the θ dependence of the vacuum energy density. The exponential factor is the action of the single instanton. The factor $\frac{1}{\lambda^8}$ arises because the integration of each zero mode parameter is accompanied by a factor $\sqrt{\frac{S_o}{2\pi\hbar}}$, with $S_o = \frac{8\pi^2}{\lambda^2}$, where λ is the coupling constant. Next, since the size ρ of the instanton has length dimension, the function f in the integrand must be dimensionless. This is achieved by setting, the argument of the function f to be proportional to $\rho\mu$, where μ is the (renormalization group) energy scale so that the product $\rho\mu$ is dimensionless. To continue, we need to determine the expression for the function f. From the running coupling, the dependence on μ of the coupling is given by, [21],

$$\frac{d\lambda}{d\log(\mu)} = -\beta_1\lambda^3, \qquad \text{where} \quad \beta_1 = \frac{11}{12\pi^2}. \tag{4.8.2}$$

Integration of this equation is straightforward. The result is

$$\lambda^2 = \frac{\lambda_o^2}{1+2\lambda_o^2\beta_1\log(\mu)}, \qquad \text{where } \lambda_o \text{ is a constant of integration.} \tag{4.8.3}$$

Now, we compute

$$\begin{aligned}
\frac{e^{-\frac{8\pi^2}{\lambda^2}}}{\lambda^8} &= \frac{e^{-\frac{8\pi^2}{\lambda_o^2}(1+2\lambda_o^2\beta_1\log(\mu))}}{\frac{\lambda_o^8}{[1+2\lambda_o^2\beta_1\log(\mu)]^4}}, \\
&= \frac{e^{-\frac{8\pi^2}{\lambda_o^2}}\mu^{-16\pi^2\beta_1}}{\lambda_o^8}\left(1+2\lambda_o^2\beta_1\log(\mu)\right)^4, \\
&\approx \frac{e^{-\frac{8\pi^2}{\lambda_o^2}}}{\lambda_o^8}\mu^{-16\pi^2\beta_1}.
\end{aligned}$$

Since (4.5.18) and (4.8.1) must not depend on μ, we read from this result the form of the function f

$$f(\rho\mu) = (\rho\mu)^{16\pi^2\beta_1}. \tag{4.8.4}$$

Hence the final expression of (4.5.18) must be of the form

$$\frac{E(\theta)}{V} = A\frac{\cos(2\pi\theta)e^{-\frac{8\pi^2}{\lambda^2}}}{\lambda^8}\int_0^\infty \frac{d\rho}{\rho^5}(\rho\mu)^{16\pi^2\beta_1}. \tag{4.8.5}$$

Now, if we plot the expression of the coupling given in (4.8.3), we will have the following profile.

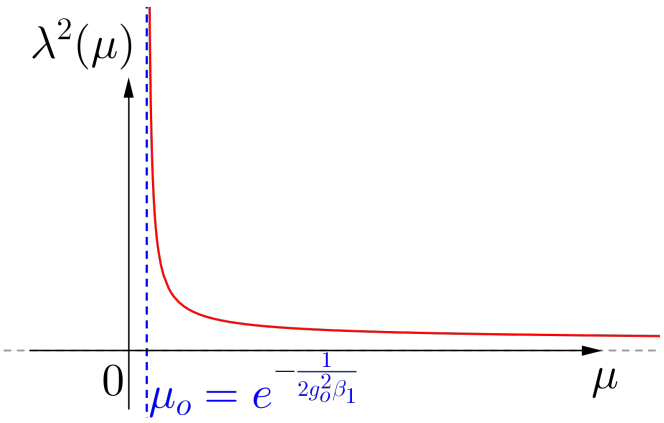


Figure 4.7: Profile of the coupling λ in terms of the energy scale μ.

We read from this figure that the theory becomes strongly coupled as we get closer to the IR. In contrast, in the UV the strength of the coupling goes to zero.

Now return to the zero mode integral, (4.8.5). Since $16\pi^2\beta_1 > 5$, the ρ integral does not converge. The divergent contribution comes from large ρ portion of the integral corresponding to instantons of a very large size. This is the IR regime of the theory. In this regime, the effective coupling is large and consequently we can't trust our semiclassical computation. This is often called the infrared problem. This is disappointing, since it prevents us from completing our semi-classical computation.

4.9 ADHM construction

We have seen the 't Hooft solution for the selfdual field strength equation, for one instanton, for the $SU(2)$ Yang-Mills theory. The clever insight in 't Hooft's ansatz is to form selfdual (or anti-selfdual) tensors of order two, from the three Pauli matrices. Based on the one instaton solution, Witten and 't Hooft were able to obtain any higher instanton number solution for gauge group $SU(2)$. Despite this success, the construction of instantons in general, for a gauge group that is not $SU(2)$, is not possible. Accordingly, one needs to find another way of solving the selfdual field equations, preferably one that does not rely heavily on the choice of the gauge group.

One method is the ADHM construction. The name ADHM stands for Atiyah, Hitchin, Drinfeld and Manin. The power of the ADHM construction is that it obtains field configuration solutions to the selfdual field equations from algebraic matrix relations plus some constraints attached to the matrices. To begin, we will review the instanton solution for the $SU(2)$ gauge group using the ADHM construction. Then, we will discuss the ADHM construction for the orthogonal gauge group $O(n)$. Following this, we will use the ADHM construction for $O(n)$ to extend the method to $SU(n)$ and $Sp(n)$ gauge groups. Finally, we will also prove the completeness of the ADHM construction by reversing the argument i.e. starting from a known selfdual field configuration we will reconstruct the matrix relations which are the starting point of the ADHM construction.

4.9.1 ADHM construction for $SU(2)$. The basic language of the ADHM construction relies on the use of a mathematical object called a "quaternion". In fact, we have already seen a quaternion when we introduced the mapping $g^{(1)}(x) = x_4 + i\vec{x}\cdot\vec{\sigma}$ as the standard mapping needed to obtain the higher winding number maps. The quantity, $x = x_4 + i\vec{x}\cdot\vec{\sigma}$, is known as a quaternion. More explicitly, for any

$(x_1, x_2, x_3, x_4) \in \mathbb{R}^4$, in terms of the Pauli matrices plus the identity matrix

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma_4 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{4.9.1}$$

a quaternion x is by definition the following 2×2 matrix

$$x = i\vec{x} \cdot \vec{\sigma} + x_4\sigma_4 = \begin{pmatrix} x_4 + ix_3 & x_2 + ix_1 \\ -x_2 + ix_1 & x_4 - ix_3 \end{pmatrix}. \tag{4.9.2}$$

A quaternion can be understood as an element of a four dimensional real vector space with basis vectors $\{\tau_\mu\}_{\mu=1,\ldots,4}$, where $\vec{\tau} = i\vec{\sigma}$ and $\tau_4 = \sigma_4 = \mathbb{1}_{2\times 2}$. Thus, we have $x = x_\mu \tau_\mu$.

Recall that the single instanton solution for the gauge field has the form

$$\mathcal{A}_\mu = \frac{r^2}{1+r^2} g^{(1)} \partial_\mu [g^{(1)}]^{-1}, \quad \text{where} \quad r = \sqrt{\textstyle\sum_{i=1}^4 x_i^2}. \tag{4.9.3}$$

We also recall that all possible mappings with different winding numbers were obtained from powers of the mapping $g^{(1)}$. Thus, it is clear that the mapping $g^{(1)}$ plays a central role in the ADHM construction. We now show that finding the solutions of the selfdual equation

$$\mathcal{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta}\mathcal{F}_{\alpha\beta}, \quad \text{where} \quad \mathcal{F}_{\mu\nu} = \partial_\mu \mathcal{A}_\nu - \partial_\nu \mathcal{A}_\mu + [\mathcal{A}_\mu, \mathcal{A}_\nu],$$

is equivalent to solving certain algebraic equations.

To begin with, define a $(1+k) \times k$ matrix $M(x)$ whose elements are quaternions. Concretely, all the entries of the matrix M are decomposed into the basis $\{\tau_\mu\}$ as

$$M_{ij} = M_{4ij}\tau_4 + M_{1ij}\tau_1 + M_{2ij}\tau_2 + M_{3ij}\tau_3. \tag{4.9.4}$$

The number k defining the dimension of the matrix M is the instanton number. The dependence of M on the quaternion x is

$$M_{ij}(x) = B_{ij} - C_{ij}x. \tag{4.9.5}$$

The matrices B and C are defined to be constant quaternion matrices. In addition the extra condition that we want the matrix M to satisfy is

$$M^\dagger M = R, \tag{4.9.6}$$

where R is required to be a real and invertible $k \times k$ matrix. Knowing that the Pauli matrices are Hermitian (so that $\tau_i|_{i=1,2,3}$ are anti-Hermitian), it is clear that the entries of the $k \times (1+k)$ dimensional $M^\dagger$ matrix are

$$M^\dagger_{ij} = M_{4ji}\tau_4 - M_{1ji}\tau_1 - M_{2ji}\tau_2 - M_{3ji}\tau_3 \tag{4.9.7}$$

To construct the selfdual gauge field $\mathcal{A}_\mu(x)$ we also need a quaternionic matrix $N(x)$ of dimension $(k+1) \times 1$. N is a quaternion column vector of dimension $k+1$. The constraints that we need N to obey are the followings

$$N^\dagger(x)M(x) = 0, \tag{4.9.8}$$

$$N^\dagger N = \mathbb{1}_{2\times 2}. \tag{4.9.9}$$

The constraint (4.9.6) implies that the rank of the matrix M must be at least k. Thus, the matrix relation in (4.9.8) admits a definite non singular solution for $N(x)$. (4.9.9) is the normalization of the solution of (4.9.8). From the definitions above, we now define the gauge field as

$$\mathcal{A}_\mu = N^\dagger \partial_\mu N. \tag{4.9.10}$$

This gauge field, a 2×2 matrix, is an element of the $\mathfrak{su}(2)$ Lie algebra. Consider for example the case where $k = 1$. This is a single instanton solution. The vector $N(x)$ turns out to be proportional to the quaternion $x = g^{(1)}$. Thus, we have

$$\mathcal{A}_\mu \propto g^{(1)} \partial_\mu [g^{(1)}]^{-1}.$$

It is also straightforward to verify that the gauge field defined in (4.9.10) is an anti-Hermitian 2×2 traceless matrix. This is indeed an element of $\mathfrak{su}(2)$.

Now, plugging the expression for the gauge field in (4.9.10) into the definition of the field strengths, we obtain

$$\begin{aligned}
\mathcal{F}_{\mu\nu} &= \partial_\mu (N^\dagger \partial_\nu N) - \partial_\nu (N^\dagger \partial_\mu N) + [N^\dagger \partial_\mu N, N^\dagger \partial_\nu N], && (4.9.11)\\
&= \partial_\mu N^\dagger \partial_\nu N - \partial_\nu N^\dagger \partial_\mu N + N^\dagger \partial_\mu N N^\dagger \partial_\nu N - N^\dagger \partial_\nu N N^\dagger \partial_\mu N, && (4.9.12)\\
&= \partial_\mu N^\dagger \left[\mathbb{1}_{(k+1)(k+1)} - N N^\dagger\right] \partial_\nu N - \partial_\nu N^\dagger \left[\mathbb{1}_{(k+1)(k+1)} - N N^\dagger\right] \partial_\mu N. && (4.9.13)
\end{aligned}$$

The passage to the last line has used the normalization condition for N. It is straightforward to see that the operator $\left[\mathbb{1}_{(k+1)(k+1)} - N N^\dagger\right]$ is the projection operator projecting to the k-dimensional quaternion subspace orthogonal to the vector N. Indeed, quick calculation shows that

$$\begin{aligned}
P &= \left[\mathbb{1}_{(k+1)(k+1)} - N N^\dagger\right], \\
&= \left[\mathbb{1}_{(k+1)(k+1)} - 2 N N^\dagger + N N^\dagger N N^\dagger\right], \quad \text{because} \quad N^\dagger N = \mathbb{1}_{2\times 2}; \\
&= \left[\mathbb{1}_{(k+1)(k+1)} - N N^\dagger\right] \left[\mathbb{1}_{(k+1)(k+1)} - N N^\dagger\right], \\
&= P^2.
\end{aligned}$$

If we return to the relation (4.9.6) we can define the projection operator $M^\dagger R^{-1} M$. Indeed, using the associativity of the matrix product we verify that $M R^{-1} M^\dagger$ is a projector

$$\begin{aligned}
(M R^{-1} M^\dagger)^2 = (M R^{-1} M^\dagger)(M R^{-1} M^\dagger) &= M R^{-1} (M^\dagger M) R^{-1} M^\dagger, \\
&= M R^{-1} R R^{-1} M^\dagger = M R^{-1} M^\dagger.
\end{aligned}$$

Now, using the relation (4.9.8), it is straightforward to see that the two projectors defined previously are equal. In other words

$$\mathbb{1}_{(1+k)(1+k)} - N N^\dagger = M R^{-1} M^\dagger. \tag{4.9.14}$$

Using this identity, the expression of the field strengths given by (4.9.13) becomes

$$\begin{aligned}
\mathcal{F}_{\mu\nu} &= \partial_\mu N^\dagger \left[\mathbb{1}_{(k+1)(k+1)} - N N^\dagger\right] \partial_\nu N - \partial_\nu N^\dagger \left[\mathbb{1}_{(k+1)(k+1)} - N N^\dagger\right] \partial_\mu N, && (4.9.15)\\
&= \partial_\mu N^\dagger M R^{-1} M^\dagger \partial_\nu N - \partial_\nu N^\dagger M R^{-1} M^\dagger \partial_\mu N, && (4.9.16)\\
&= N^\dagger \left(\partial_\mu M R^{-1} \partial_\nu M^\dagger - \partial_\nu M R^{-1} \partial_\mu M^\dagger\right) N, && (4.9.17)\\
&= N^\dagger C \left(\tau_\mu R^{-1} \tau_\nu^\dagger - \tau_\nu R^{-1} \tau_\mu^\dagger\right) C^\dagger N. && (4.9.18)
\end{aligned}$$

The passage to the third line has used the identity $\partial_\mu N^\dagger M = -N^\dagger \partial_\mu M$ while the passage to the last line has used the alternative definition of M in (4.9.5). Recall that the matrix R is real, so that it commutes with the unit quaternion τ_μ. Hence the final expression of the field strength is

$$\mathcal{F}_{\mu\nu} = N^\dagger C R^{-1}(\tau_\mu \tau_\nu^\dagger - \tau_\nu \tau_\mu^\dagger) C^\dagger N. \tag{4.9.19}$$

It is straightforward to check that $(\tau_\mu \tau_\nu^\dagger - \tau_\nu \tau_\mu^\dagger)$ is selfdual and satisfies the relation

$$\tau_\mu \tau_\nu^\dagger - \tau_\nu \tau_\mu^\dagger = i4\sigma_{\mu\nu},$$

where

$$\sigma_{\mu\nu} = \begin{cases} -\frac{i}{4}[\sigma_i, \sigma_j] & \mu = i, \quad \nu = j \\ \frac{1}{2}\sigma_i & \mu = i, \quad \nu = 4 \end{cases}$$

is the selfdual antisymmetric tensor defined in the ansatz of 't Hooft single instanton solution for $SU(2)$. Accordingly, the final compact form of the selfdual field strength in terms of $\sigma_{\mu\nu}$ is

$$\mathcal{F}_{\mu\nu} = i4 N^\dagger C R^{-1} \sigma_{\mu\nu} C^\dagger N. \tag{4.9.20}$$

This proves that solving the selfdual field equation is equivalent to solving the matrix equation (4.9.8) with constraints on the matrices N and M. Notice that the only gauge group dependence of the matrices N and M appears in the dimension of the matrices. This observation will lead us to the ADHM construction for $O(n)$ gauge group.

4.9.2 ADHM construction for $O(n)$. The central ingredients of the ADHM construction are the matrices $M(x)$ and $N(x)$. For $O(n)$ the matrix $M(x)$ is defined to be a $(2k+n)\times k$ matrix with quaternion entries. Again the number k appearing in the dimension of the matrix is the instanton number. The definitions and constraints associated to the matrix $M(x)$ that we gave above are unchanged. The only change is that the dimension of the matrix is now $(2k+n)\times k$. In particular, we still have the two constant quaternion matrices B and C such that

$$[M(x)]_{ij} = B_{ij} - C_{ij}x, \quad \text{where} \quad \begin{cases} 1 \le i \le 2k+n, \\ 1 \le j \le k. \end{cases} \tag{4.9.21}$$

We still require that the rank of the matrix M is k, so that the matrix R defined as

$$\Re\left(\sum_{l=1}^{2k+n} M_{li}^\dagger q M_{lj}\right) \equiv R_{ij}, \quad \text{where } q \text{ is a constant quaternion,} \tag{4.9.22}$$

is invertible for any quaternion $q \neq 0$. $\Re$ denotes the real part of the complex number in the entries of the matrix. By definition R_{ij} is a real invertible matrix. Next, we define n real $(2k+n)-$dimensional vectors $N_i|_{1\le i\le n}$. In fact, we can visualize N as a $(2k+n)\times n$ matrix. The extension to $O(n)$ of the condition given in (4.9.9) is

$$N_i^T N_j = \sum_{l=1}^{2k+n} N_{li}^T N_{lj} = \delta_{ij}. \tag{4.9.23}$$

N_i^T is the row vector transpose of the N_i column vector. The additional constraints we impose on M and N are

$$\sum_{l=1}^{2k+n} N_{li}^T M_{lj} = 0. \tag{4.9.24}$$

Again, this relation is the natural extension to $O(n)$ of the condition (4.9.8). In summary, the above equations can be written as

$$N^T M = 0, \tag{4.9.25}$$
$$N^T N = \mathbb{1}_{n\times n}. \tag{4.9.26}$$

The requirement that N_i is a real vector follows because the Lie algebra $\mathfrak{o}(n)$ is a real vector space, unlike the case of $\mathfrak{su}(2)$. Ultimately this is also why we defined R_{ij} to be a real matrix. Following the same steps as above, we now define the gauge field

$$\mathcal{A}_\mu = N^T \partial_\mu N. \tag{4.9.27}$$

It is straightforward to see that $\mathcal{A}_\mu$ is indeed an element of $\mathfrak{o}(n)$. Arguing as we did above we find that the expression for the field strength is

$$\mathcal{F}_{\mu\nu} = \partial_\mu N^T \left(\mathbb{1}_{(2k+n)\times(2k+n)} - NN^T\right) \partial_\nu N - \partial_\nu N^T \left(\mathbb{1}_{(2k+n)\times(2k+n)} - NN^T\right) \partial_\mu N. \tag{4.9.28}$$

Again, we can use the condition (4.9.25) to show that

$$\mathbb{1}_{(2k+n)\times(2k+n)} - NN^T = \Re(M^\dagger R^{-1} M).$$

Thus, we can carry out the same calculations as in (4.9.18) to end up with the following expression for the field strengths

$$\mathcal{F}_{\mu\nu} = N^T \Re\left[C R^{-1}\left(\tau_\mu \tau_\nu^\dagger - \tau_\nu \tau_\mu^\dagger\right) C^\dagger\right] N, \tag{4.9.29}$$
$$= 4 N^T \Re\left(i C R^{-1} \sigma_{\mu\nu} C^\dagger\right) N. \tag{4.9.30}$$

This completes our discussion of the ADHM construction for $O(n)$.

4.9.3 ADHM construction for $SU(n)$ and $Sp(n)$. The ADHM construction for gauge group $SU(n)$ and $Sp(n)$ is easily constructed once we have carried out the construction for the orthogonal group $O(n)$. In fact, we can realize the $SU(n)$ group as a subgroup of the orthogonal $O(2n)$ group. Thus, the solutions found for $O(2n)$ can, with minor changes, yield solutions for $SU(n)$. The case of $Sp(n)$ is very similar since it can also be realized as a subgroup of $O(4n)$.

Since we have already set up all the ingredients of the ADHM construction for $O(n)$, we will skip common calculations in the next paragraphs. We begin with the $SU(n)$ construction. The extra constraint that we need to impose is that the real $2n \times 2n$ matrix $\mathcal{A}_\mu \in \mathfrak{o}(2n)$ acts as an $n \times n$ anti-Hermitian $\tilde{\mathcal{A}}_\mu \in \mathfrak{su}(n)$ when acting on an $n-$dimensional complex plane. Thus the $2n$ real dimensions where $\mathcal{A}_\mu$ acts, are identified as the real and imaginary parts of $\tilde{\mathcal{A}}_\mu$. Accordingly, the dimension of the quaternionic matrix $M(x)$ must be $(4k+2n)\times k$. In addition, for the group $SU(n)$, we require

$$JM(x) = \tau_1 \mathbb{1}_{(4k+2n)\times(4k+2n)} M(x), \quad \text{where} \quad J = \begin{pmatrix} 0 & \mathbb{1}_{(2k+n)\times(2k+n)} \\ -\mathbb{1}_{(2k+n)\times(2k+n)} & 0 \end{pmatrix}. \tag{4.9.31}$$

This extra condition for $SU(n)$ is equivalent to choosing $M(x)$ such that

$$M(x) = \begin{pmatrix} \tilde{M}(x) \\ \tau_1 \tilde{M}(x) \end{pmatrix}, \tag{4.9.32}$$

where $\tilde{M}$ is a $(2k+n)\times k$ quaternionic matrix. We now recover the same conditions as for the $O(n)$ construction. Consequently the reality and invertibility condition is written as

$$\Re\left[\tilde{M}^\dagger(q-\tau_1 q\tau_1)\tilde{M}\right]_{ij}=R_{ij}. \tag{4.9.33}$$

Note that the matrix R on the left hand side must be a square $k\times k$ invertible matrix. Next, the appropriate dimension for the n vectors N_i is $(4k+2n)$. If we work with the tilde vectors, we also need to split each N_i vector into two blocks of column vectors $\tilde{N}_i$ of size $(2k+n)$. It is straightforward to show that the relation between the components of tilde vectors and the original vector must satisfy

$$\tilde{N}_{ij}=N_{ij}-N_{(i+2k+n)j},\quad \begin{cases}1\le i\le 2k+n\\ \le j\le n.\end{cases} \tag{4.9.34}$$

Now, the algebraic constraints are

$$\tilde{N}^\dagger\tilde{M}=0, \tag{4.9.35}$$
$$\tilde{N}^\dagger\tilde{N}=\mathbb{1}_{n\times n}, \tag{4.9.36}$$

and the gauge field is

$$\tilde{\mathcal{A}}_\mu=\tilde{N}^\dagger\partial_\mu\tilde{N}. \tag{4.9.37}$$

The computation of the projector operator and the field strengths in terms of the matrix $\tilde{M}$ are easy to derive using the above matrix relations.

The ADHM construction for the $Sp(n)$ group basically uses the same trick as for $SU(n)$. The difference is that for $Sp(n)$ we introduce three $(4k+4n)\times(4k+4n)$ matrices $J_i|_{i=1,2,3}$ such that

$$J_i^2=-\mathbb{1}_{(4k+4n)\times(4k+4n)},\quad J_1J_2=J_3. \tag{4.9.38}$$

Hence, we define the matrix $M(x)$ of $(4k+4n)\times k$ dimensions satisfying

$$J_1M=\tau_1M,\quad J_2M=\tau_2M. \tag{4.9.39}$$

If we write the J_i matrices in terms of $(k+n)\times(k+n)$ blocks, we can check that the following matrices satisfy the algebra given in (4.9.38)

$$J_1=\begin{pmatrix}0 & \mathbb{1}_{(n+k)\times(n+k)} & 0 & 0\\ -\mathbb{1}_{(n+k)\times(n+k)} & 0 & 0 & 0\\ 0 & 0 & 0 & \mathbb{1}_{(n+k)\times(n+k)}\\ 0 & 0 & -\mathbb{1}_{(n+k)\times(n+k)} & 0\end{pmatrix},$$
$$J_2=\begin{pmatrix}0 & 0 & \mathbb{1}_{(n+k)\times(n+k)} & 0\\ 0 & 0 & 0 & -\mathbb{1}_{(n+k)\times(n+k)}\\ -\mathbb{1}_{(n+k)\times(n+k)} & 0 & 0 & 0\\ 0 & \mathbb{1}_{(n+k)\times(n+k)} & 0 & 0\end{pmatrix},$$
$$J_3=\begin{pmatrix}0 & 0 & 0 & -\mathbb{1}_{(n+k)\times(n+k)}\\ 0 & 0 & -\mathbb{1}_{(n+k)\times(n+k)} & 0\\ 0 & \mathbb{1}_{(n+k)\times(n+k)} & 0 & 0\\ \mathbb{1}_{(n+k)\times(n+k)} & 0 & 0 & 0\end{pmatrix}.$$

Hence the matrix M can be written in terms of a single $(k+n)\times k$ matrix $\tilde{M}$ as

$$M(x) = \begin{pmatrix} \tilde{M} \\ \tau_1\tilde{M} \\ \tau_2\tilde{M} \\ \tau_3\tilde{M} \end{pmatrix} \tag{4.9.40}$$

The reality and invertibility conditions become

$$\Re\left[\tilde{M}^\dagger(q - q\tau_1 q - q\tau_2 q - q\tau_3 q)\tilde{M}\right]_{ij} = R_{ij}. \tag{4.9.41}$$

Again, we need to split the vectors N_i into blocks of four $n+k$ columns vectors $\tilde{N}_i$. Thus, with the ADHM constraints imposed on N_i and M, the gauge field solution to the selfdual field equation is

$$\tilde{A}_\mu = \tilde{N}^\dagger\partial_\mu\tilde{N}. \tag{4.9.42}$$

The derivation of the projectors and the fields strengths are straightforward. In fact, based on our previous detailed calculations, the field strengths will always have the following form

$$\mathcal{F}_{\mu\nu} = i4\tilde{N}^\dagger\tilde{C}R^{-1}\sigma_{\mu\nu}\tilde{C}^\dagger\tilde{N}. \tag{4.9.43}$$

There is a gap in the logic of our discussion: it is not clear that the number k appearing in the dimension of the ADHM matrices is the instanton number. To complete our discussion we will prove that k is indeed the winding number for the ADHM construction of the selfdual field. To begin with, recall that the formula for the winding number in terms of the field strength is

$$k = -\frac{1}{32\pi^2}\int d^4x \mathrm{tr}\left(\epsilon_{\mu\nu\alpha\beta}\mathcal{F}_{\mu\nu}\mathcal{F}_{\alpha\beta}\right). \tag{4.9.44}$$

Next, consider the expression of the selfdual field strength given by the ADHM construction.

$$\mathcal{F}_{\mu\nu} \propto \tilde{N}^\dagger\tilde{C}R^{-1}\sigma_{\mu\nu}\tilde{C}^\dagger\tilde{N},$$

where (for example for the case of the $Sp(n)$ group) the dimension of the matrices $\tilde{N}$, $\tilde{C}$ and $\tilde{R}$ are respectively $(n+k)\times n$, $(n+k)\times k$ and $k\times k$. To simplify our calculation we discuss the $Sp(n)$ gauge group for concreteness. The main idea of the proof is the same for the other gauge groups except that we need to account for the appropriate dimensions of the above three matrices. The dimension of the matrix R is always fixed to $k\times k$ in our case.

Proof. We start by recalling one of the most important constraints on the matrix M of the ADHM construction: it must be of rank k. Therefore, there must be some orthogonal matrices $S\in O(n+k)$ and $T\in O(k)$ transforming

$$\begin{aligned}
\tilde{M} &\longrightarrow \tilde{M}' = \tilde{B}' + \tilde{C}'x = S\tilde{M}T,\\
\tilde{N} &\longrightarrow \tilde{N}' = S\tilde{N},\\
R &\longrightarrow R' = T^T R T,
\end{aligned}$$

without changing the selfdual field solution. The expression for the field strength in terms of these new matrices becomes

$$\begin{aligned}
\mathcal{F}_{\mu\nu} \longrightarrow \mathcal{F}'_{\mu\nu} &= i4\tilde{N}'^\dagger\tilde{C}'R'^{-1}\sigma_{\mu\nu}\tilde{C}'^\dagger\tilde{N}',\\
&= i4\tilde{N}^\dagger S^T S\tilde{C}TT^T R^{-1}T\sigma_{\mu\nu}T^T\tilde{C}^\dagger S^T S\tilde{N},\\
&= i4\tilde{N}^\dagger\tilde{C}R^{-1}\sigma_{\mu\nu}\tilde{C}^\dagger\tilde{N},\\
&= \mathcal{F}_{\mu\nu}.
\end{aligned}$$

The passage to the last line has used the above three transformation laws and the fact that S and T are orthogonal matrices. This last equation shows that the ADHM construction is indeed invariant under these transformations. It is straightforward to see that there exist matrices S and T, keeping the rank of $\tilde{M}'$ to be k, such that

$$\tilde{C}' = \begin{pmatrix} \left.\begin{pmatrix} \overbrace{\begin{matrix} 0 & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 \end{matrix}}^{k} \end{pmatrix}\right\} n \\ \left.\begin{pmatrix} c_{(n+1)1} & c_{(n+1)2} & \cdots & c_{(n+1)k} \\ 0 & c_{(n+2)2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \underbrace{0 \quad \cdots \quad 0}_{k} & & & c_{(n+k)k} \end{pmatrix}\right\} k \end{pmatrix},$$

where the k constant and quaternionic column vectors $c_j|_{j=1,\cdots,k} = c_{ij}|_{j=1,\cdots,k;\ n+1\leq i\leq n+k}$ are linearly independent. Now, we can see from the matrix product $\tilde{N}'^{\dagger}\tilde{C}'$ that the only relevant components of the vector $\tilde{N}'$ and thus of $\tilde{N}$ are those below the n^{th} lines. Thus, there are exactly $n+k-n=k$ copies of the quaternion x (because N is the only matrix depending on x) distributed among the diagonal entries of the matrix $\mathcal{F}_{\mu\nu}$. A detailed computation now shows, [22]

$$\begin{aligned} \text{tr}\,(\mathcal{F}_{\mu\nu}\mathcal{F}_{\mu\nu}) &= -16\text{tr}\left[(\tilde{N}^{\dagger}\tilde{C}R^{-1}\sigma_{\mu\nu}\tilde{C}^{\dagger}\tilde{N})(\tilde{N}^{\dagger}\tilde{C}R^{-1}\sigma_{\mu\nu}\tilde{C}^{\dagger}\tilde{N})\right], && (4.9.45)\\ &= \alpha(x_\mu)k. && (4.9.46)\end{aligned}$$

From the explicit form for $\alpha(x_\mu)$ it is possible to show that

$$\int d^4x\alpha(x_\mu) = -16\pi^2. \tag{4.9.47}$$

Hence, the number k appearing in the ADHM construction matrices is indeed the instanton number. □

4.10 The completeness of the ADHM construction

To prove that all possible instanton solutions are obtained from the ADHM construction, we need to show that:

- Given any selfdual (or anti-selfdual) gauge potential $\mathcal{A}_\mu$, we can always construct (up to gauge transformations) the matrices appearing in the ADHM construction.
- Given these matrices, we can reproduce the original potential we started with.

The proof is rather technical so we will only summarize the logic of the arguments presented by Corrigan and Goddard [23]. They pointed out that massless Dirac fields are sensitive to the presence of instantons. They carry information about the instanton field itself and this is the reason why we are studying these fields. To this end, it is natural to find the solution of the massless Dirac equation in the background of the instanton field. Indeed, by finding the solution of the massless Dirac equation in the background of

the vector potential one can show that a given selfdual field configuration implies some algebraic matrix equations. The matrix relations are the algebraic constraints that one needs to set up the ADHM construction.

Recall that the covariant derivative in Euclidean space and the Dirac matrices are respectively

$$D_\alpha = \partial_\alpha + \mathcal{A}_\alpha, \quad \gamma_\alpha = \begin{pmatrix} 0 & \tau_\alpha \\ \tau_\alpha^\dagger & 0 \end{pmatrix}. \tag{4.10.1}$$

It is straightforward to check that the Dirac matrices satisfy the anticommutation relation

$$\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu}\mathbb{1}_{4\times 4}. \tag{4.10.2}$$

The massless Dirac equation is now defined by

$$\gamma_\mu D_\mu \Psi = 0, \tag{4.10.3}$$

where Ψ is a four component massless spinor. Defining the projectors

$$P_+ = \frac{1}{2}\left(\mathbb{1}_{4\times 4} + \gamma_5\right),$$
$$P_- = \frac{1}{2}\left(\mathbb{1}_{4\times 4} - \gamma_5\right), \quad \text{where} \quad \gamma_5 = \gamma_1\gamma_2\gamma_3\gamma_4 = \begin{pmatrix} -\mathbb{1}_{2\times 2} & 0 \\ 0 & \mathbb{1}_{2\times 2} \end{pmatrix},$$

we can split the spinor Ψ in two spinors Ψ_+ and Ψ_- containing the components of the positive and negative chirality

$$\Psi = \begin{pmatrix} \Psi_- \\ \Psi_+ \end{pmatrix}$$

Accordingly Dirac's equation becomes

$$\tau_\mu^\dagger D_\mu \Psi_- = 0, \qquad \tau_\nu D_\nu \Psi_+ = 0. \tag{4.10.4}$$

Now, it is straightforward to show by direct calculation that

$$\tau_\mu^\dagger \tau_\nu = \delta_{\mu\nu} + i2\bar{\sigma}_{\mu\nu}, \tag{4.10.5}$$
$$\tau_\mu \tau_\nu^\dagger = \delta_{\mu\nu} + i2\sigma_{\mu\nu}. \tag{4.10.6}$$

where

$$\bar{\sigma}_{\mu\nu} = \begin{cases} \bar{\sigma}_{ij} = \sigma_{ij}, \\ \bar{\sigma}_{i4} = -\sigma_{i4} \end{cases}$$

is the anti-selfdual tensor defined in 't Hooft's single instanton solution. Accordingly, we can use the above results and compute

$$\tau_\mu^\dagger D_\mu \tau_\nu D_\nu = \tau_\mu^\dagger \tau_\nu D_\mu D_\nu, \tag{4.10.7}$$
$$= (\delta_{\mu\nu} + i2\bar{\sigma}_{\mu\nu}) D_\mu D_\nu, \tag{4.10.8}$$
$$= D^2 + i\bar{\sigma}_{\mu\nu}[D_\mu, D_\nu], \tag{4.10.9}$$
$$= D^2 + i\bar{\sigma}_{\mu\nu}\mathcal{F}_{\mu\nu}. \tag{4.10.10}$$

The passage to the last line has used the fact that $\bar{\sigma}_{\mu\nu}$ is antisymmetric under swapping $\mu \leftrightarrow \nu$. Since $\mathcal{F}_{\mu\nu}$ is selfdual and $\bar{\sigma}_{\mu\nu}$ is anti-selfdual, the combination $\bar{\sigma}_{\mu\nu}\mathcal{F}_{\mu\nu}$ has to satisfy

$$\bar{\sigma}_{\mu\nu}\mathcal{F}_{\mu\nu} = 0. \tag{4.10.11}$$

Now, we use the equation (4.10.4) to obtain

$$D^2\Psi_+ = 0. \tag{4.10.12}$$

This equation involves only the positive chirality component of the spinor. This equation does not admit a normalizable solution. Hence the massless Dirac equation has no normalizable positive chirality solutions. However, an application of an index theorem, [24, 25], shows that it has k linearly independent normalizable negative chirality solutions, Ψ_-. Accordingly, the solution Ψ_- can be put into the form of a $2n+k$ dimensional matrix ψ with the elements ψ_{iaI}, where $1 \leq i \leq k, \quad 1 \leq a \leq n, \quad I = 1, 2$. Following the argument of Corrigan and Goddard, we take the transpose of ψ_{iaI} in the two-dimensional label and then multiply it by

$$\epsilon = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{4.10.13}$$

to obtain an $n \times 2k$ dimensional matrix $\tilde{\psi} = \psi^T\epsilon$. In order to obtain the ADHM matrices, we need to integrate the formula [26]

$$\psi^\dagger\psi = -\frac{1}{4}\partial^2 R^{-1}, \tag{4.10.14}$$

where R is a $k \times k$ real and invertible matrix. After integration and performing the above change $\psi \to \tilde{\psi}$, we obtain the solution, [23, 27],

$$\tilde{\psi} = N^\dagger C R, \tag{4.10.15}$$

where $N, \quad C$ are $(2k+n) \times n$, $(2k+n) \times k$ dimensional matrices respectively. In the solution above N, C and R are the ADHM matrices. Details of the previous calculation can be found in [23, 27]. The only matrix left to complete the ADHM construction is the matrix B. One way of finding the matrix B is to study the large x behaviour of ψ. Since $M(x) \underset{|x|\to\infty}{\longrightarrow} Cx$ it is shown in [23] that the solution for $\tilde{\psi}$ has the following form for large $|x|$

$$\tilde{\psi} = -gN^\dagger B\frac{x^\dagger}{|x|^4}, \quad \text{where } g \in SU(n). \tag{4.10.16}$$

Finally, we conclude that, starting from the $SU(n)$ selfdual instanton solutions, one can show that $M = B + Cx$ and $N(x)$ can be reconstructed, up to gauge transformation. The trick in finding these matrices strongly relies on the massless Dirac equation in the background of the field $\mathcal{A}_\mu$.

4.11 The moduli space of an instanton

When we talk about the moduli space of an instanton, we mean the space of all inequivalent instanton solutions. As a concrete example, the moduli space of an instanton, denoted $\mathbb{M}_{1,1}$ in one dimensional quantum mechanics is the trivial topological space $\mathbb{R}$. This is because different instanton solutions correspond to different values of the center of the instanton parameter, and the instanton center can vary in the real line. In general, instanton moduli spaces are manifolds with each point in the manifold corresponding to a particular solution.

When we come to discuss Yang-Mills theory, we can further refine the notion of the moduli space by considering the space of all instanton configurations that share a given instanton number k. In the case of $SU(2)$ Euclidean Yang-Mills theory, the dimension of the manifold isomorphic to the moduli space of instantons, denoted $\mathbb{M}_{k,2}$, is $D = 8k$, where k is the instanton number. This follows because there are $8k$ arbitrary parameters appearing in the k instanton solutions to the selfdual field equations. Since all gauge transformations of the solutions are physically equivalent, the moduli space of instantons is quotiented by local gauge transformations. Of course, as we described above the moduli space must only include solutions that differ by a global gauge transformation not solutions that differ by a local gauge transformation.

Since a moduli space is a manifold, it is possible to define intrinsic and geometrical properties of the moduli space. Indeed, if we consider for example the $U(N)$ gauge group, we know that the gauge field $\mathcal{A}_\mu$ is an element of the $\mathfrak{u}(N)$ Lie algebra and thus we can generalize the inner product found in (3.9.20) as follows

$$(\delta\mathcal{A}_\mu, \delta\mathcal{A}'_\nu) = \int d^4x \mathrm{tr}(\delta\mathcal{A}_\mu \delta\mathcal{A}_\nu)\delta_{\mu\nu}. \tag{4.11.1}$$

In this way we see that the moduli space denoted $\mathbb{M}_{k,N}$ for k instantons is endowed with a natural metric. This metric written in (4.11.1) is not quite well defined, as we have not yet quotiented out by local gauge transformations. Below we will return to this point and will define a metric for moduli space that is gauge invariant.

Now, fix the instanton number to k and consider the moduli space for all selfdual field solutions. If $\mathcal{A}_\mu$ belongs to the moduli space $\mathbb{M}_{k,N}$ a gauge transformation of $\mathcal{A}_\mu$, $\mathcal{A}_\mu + \delta\mathcal{A}_\mu$ must also be a solution and thus, it too belongs to $\mathbb{M}_{k,N}$. The corresponding gauge transformation of the covariant derivative is $D_\mu \to \partial_\mu + \mathcal{A}_\mu + \delta\mathcal{A}_\mu$. The linearized selfdual field equations imply that

$$D_\mu(\delta\mathcal{A}_\nu) - D_\nu(\delta\mathcal{A}_\mu) = \epsilon_{\mu\nu\rho\sigma} D_\rho(\delta\mathcal{A}_\sigma). \tag{4.11.2}$$

Indeed, a straightforward calculation shows that

$$\begin{aligned}
[D_\mu + \delta\mathcal{A}_\mu, D_\nu + \delta\mathcal{A}_\nu]\phi &= [D_\mu, D_\nu]\phi + [D_\mu, \delta\mathcal{A}_\nu]\phi + [\delta\mathcal{A}_\mu, D_\nu]\phi + O(\delta\mathcal{A}^2), \\
&= \mathcal{F}_{\mu\nu}\phi + D_\mu(\delta\mathcal{A}_\nu\phi) - \delta\mathcal{A}_\nu D_\mu\phi + D_\nu(\delta\mathcal{A}_\mu\phi) - \delta\mathcal{A}_\nu D_\mu\phi, \\
&= [\mathcal{F}_{\mu\nu} + D_\mu(\delta\mathcal{A}_\nu) - D_\nu(\delta\mathcal{A}_\mu)]\,\phi.
\end{aligned}$$

Now, use the selfdual condition $\mathcal{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\mathcal{F}_{\rho\sigma}$ for both the old and the gauge transformed field strengths to obtain

$$\begin{aligned}
D_\mu(\delta\mathcal{A}_\nu) - D_\nu(\delta\mathcal{A}_\mu) &= \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\big(D_\rho(\delta\mathcal{A}_\sigma) - D_\sigma(\delta\mathcal{A}_\rho)\big), \\
&= \epsilon_{\mu\nu\rho\sigma}\big(D_\rho(\delta\mathcal{A}_\sigma).
\end{aligned}$$

The passage to the last line has used the fact that the tensor $\epsilon_{\mu\nu\rho\sigma}$ is totally antisymmetric and we have relabeled the contracted indices.

Now, in terms of the small displacement $\delta\mathcal{A}_\mu$, we can define the metric on the moduli space as

$$g(\delta\mathcal{A}_\mu, \delta\mathcal{A}'_\nu) = \int d^4x \mathrm{tr}(\delta\mathcal{A}_\mu \delta\mathcal{A}'_\nu)\delta_{\mu\nu}. \tag{4.11.3}$$

We must require that all gauge equivalent gauge fields are collinear with their representative. One way in which we can accomplish this, is to use the metric (4.11.1) but with the extra input that the metric

is applied to tangent vectors that are first projected to be orthogonal to gauge transformations. By requiring this we are simply making a choice of representatives for tangent vectors in the moduli space after we have quotiented by local gauge transformations. By definition, our representatives obey

$$g(\delta\mathcal{A}_\mu, D_\mu\eta) = \int d^4x \mathrm{tr}(\delta\mathcal{A}_\mu D_\mu\eta) = 0, \tag{4.11.4}$$

$$= -\int d^4x \mathrm{tr}\big(D_\mu(\delta\mathcal{A}_\mu)\eta\big) = 0, \tag{4.11.5}$$

for all η. The passage to the second line has used an integration by parts. This last equation is also equivalent to

$$D_\mu\delta\mathcal{A}_\mu = 0. \tag{4.11.6}$$

Zero modes for the k instantons are solutions of the equations (4.11.2) and (4.11.6). They form the tangent space of the manifold $\mathbb{M}_{k,N}$, which is the moduli space of instantons after gauge fixing. In technical terms, the tangent bundle of the manifold $\mathbb{M}_{k,N}$ constitutes the space of the zero modes.

Now, return to the $SU(2)$ gauge group. Our goal is to obtain the metric of $\mathbb{M}_{1,2}$. Our discussion will follow the discussion given in [3]. Recall that there are $8k$ parameters appearing in the k instanton solution. We can thus define collective coordinates ζ_α on the moduli space $\mathbb{M}_{k,2}$ where α runs from 1 to $8k$. Our logic should be familiar. Indeed, recall that in chapter 2 the zero mode for the double well potential problem had the form $\frac{d\tilde{x}}{dt}$, where $\tilde{x}$ was the instanton solution. By analogy, the zero mode solutions to (4.11.2) and (4.11.6) are $\delta_\alpha\mathcal{A}_\mu = \frac{\partial A_\mu}{\partial\zeta_\alpha}$. For k = 1, the solution written to include all 8 parameters is

$$\mathcal{A}_\mu = \frac{\rho^2(x-X)_\nu}{(x-X)^2[(x-X)^2+\rho^2]} g\sigma_{\mu\nu}g^{-1}. \tag{4.11.7}$$

We can read from the above equation that the collective coordinates are $\zeta_\alpha = (X_1, X_2, X_3, X_4, \rho, \theta_1, \theta_2, \theta_3)$, where X_μ, ρ and θ_i are respectively labeling translations, scaling and global gauge directions. The solution given in (4.11.7) does not explicitly show that $\mathcal{A}_\mu$ lives in $\mathfrak{u}(2)$. However, use of the definition of $\sigma_{\mu\nu}$ shows that it can be written as $\sigma_{\mu\nu} = \eta_{i\mu\nu}\frac{\sigma_i}{2}$. Indeed

$$\sigma_{\mu\nu} = \begin{cases} -\frac{i}{4}[\sigma_i,\sigma_j] & \text{if } \mu=i,\nu=j \\ \frac{1}{2}\sigma_i & \text{if } \mu=i,\nu=4, \end{cases}$$

$$\Rightarrow \sigma_{\mu\nu} = \begin{pmatrix} 0 & \frac{1}{2}\sigma_3 & -\frac{1}{2}\sigma_2 & \frac{1}{2}\sigma_1 \\ -\frac{1}{2}\sigma_3 & 0 & \frac{1}{2}\sigma_1 & \frac{1}{2}\sigma_2 \\ \frac{1}{2}\sigma_2 & -\frac{1}{2}\sigma_1 & 0 & \frac{1}{2}\sigma_3 \\ -\frac{1}{2}\sigma_1 & -\frac{1}{2}\sigma_2 & -\frac{1}{2}\sigma_3 & 0 \end{pmatrix}_{\mu\nu}. \tag{4.11.8}$$

Now, we read the $\mathfrak{u}(2)$ components of $\sigma_{\mu\nu} = \eta_{i\mu\nu}\frac{\sigma_i}{2}$ as

$$\eta_1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}, \quad \eta_2 = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \quad \eta_3 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}. \tag{4.11.9}$$

Since only the fourth components ($\mu = i, \nu = 4$) pick up an extra minus sign in the definition of the anti-selfdual tensor $\bar{\sigma}_{\mu\nu}$, it is straightforward to see that the $\mathfrak{u}(2)$ components for the anti-selfdual matrices are

$$\bar{\eta}_1 = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad \bar{\eta}_2 = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad \bar{\eta}_3 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{4.11.10}$$

Now, using the above matrices, (4.11.7) is rewritten as

$$\mathcal{A}_\mu = \frac{1}{2}\frac{\rho^2(x-X)_\nu}{(x-X)^2[(x-X)^2+\rho^2]}\eta_{i\mu\nu}g\sigma_i g^{-1}. \tag{4.11.11}$$

This expression explicitly demonstrates that $\mathcal{A}_\mu$ is an element of $\mathfrak{u}(2)$. The translation (zero) modes up, to gauge transformation, are written as

$$\delta_{(\nu)}\mathcal{A}_\mu = \frac{\partial \mathcal{A}_\mu}{\partial X_\nu} + D_\mu \Omega_\nu, \tag{4.11.12}$$

where Ω_ν must be chosen so that (4.11.6) is satisfied. Since $\frac{\partial}{\partial x_\nu}$ and $\frac{\partial}{\partial X_\nu}$ act equivalently, it is straightforward to see that the choice $\Omega_\nu = \mathcal{A}_\nu$ is valid. Indeed, under this choice, the expression (4.11.12) yields $\delta_{(\nu)}\mathcal{A}_\mu = \mathcal{F}_{\mu\nu}$. Moreover, (4.11.6) is automatically satisfied due to the equations of motion $D_\mu\mathcal{F}_{\mu\nu} = \partial_\mu\mathcal{A}_\nu - \partial_\nu\mathcal{A}_\mu + [\mathcal{A}_\mu, \mathcal{A}_\nu] = 0$ in the background of the selfdual field configuration. Now, use (4.11.3) to obtain the component of the metric with respect to the directions $\zeta_\mu = (X_1, X_2, X_3, X_4, 0, 0, 0, 0)$ of the moduli space

$$g_{\alpha\beta}|_{\alpha,\beta\in\{1,2,3,4\}} = \int d^4x \mathrm{tr}\left[\delta_{(\alpha)}\mathcal{A}_\mu \delta_{(\beta)}\mathcal{A}_\mu\right], \tag{4.11.13}$$

$$= \int d^4x \mathrm{tr}(\mathcal{F}_{\mu\alpha}\mathcal{F}_{\mu\beta}) = 4S_o\delta_{\alpha\beta}, \tag{4.11.14}$$

$$= 32\pi^2\delta_{\alpha\beta}. \tag{4.11.15}$$

Next, we want to find the components of the metric in the directions $\zeta_\alpha = (0,0,0,0,\rho,0,0,0)$ and $\zeta_\beta = (0,0,0,0,0,\theta_1,\theta_2,\theta_3)$. These are easily obtained if we know the scaling zero mode and the global gauge orientation zero modes. Thus, we need to find $\delta\mathcal{A}_\mu = \frac{\partial\mathcal{A}_\mu}{\partial\rho}$ and $\delta_{(i)}\mathcal{A}_\mu = \frac{\partial\mathcal{A}_\mu}{\partial\theta_i}$. However, since $\mathcal{A}_\mu$ given in (4.11.11) is by definition the solution of (4.11.6), one can show, [3], that the remaining components of the metric are respectively

$$g_{\alpha\beta}|_{\alpha,\beta\in\{5\}} = \int d^4x \mathrm{tr}(\delta\mathcal{A}_\alpha\delta\mathcal{A}_\beta) = 64\pi^2\delta_{\alpha\beta}, \tag{4.11.16}$$

$$g_{\alpha\beta}|_{\alpha,\beta\in\{6,7,8\}} = \int d^4x \mathrm{tr}(\delta\mathcal{A}_\alpha\delta\mathcal{A}_\beta) = 64\pi^2\delta_{\alpha\beta}\rho^2. \tag{4.11.17}$$

Concretely the metric of the moduli space $\mathbb{M}_{1,2}$ is

$$g_{\alpha\beta} = 32\pi^2 \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2\rho^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2\rho^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2\rho^2 \end{pmatrix}_{\alpha\beta}. \tag{4.11.18}$$

Having this metric, we are tempted to write the isomorphism

$$\mathbb{M}_{1,2} \sim \mathbb{R}^4 \times \mathbb{R}^4. \tag{4.11.19}$$

However, this is not true because the fields $\mathcal{A}_\mu$ are left invariant under the center of the gauge group $\mathbb{Z}_2$. The metric obtained from the gauge rotation zero modes plus the scale zero mode is a metric on

$$\mathbb{R}^+ \times S^3/\mathbb{Z}_2 \sim \mathbb{R}^4/\mathbb{Z}_2.$$

ρ^2 belongs to $\mathbb{R}^+$, which is the set of real positive numbers. The parameters θ_i of the gauge group belong to the three sphere S^3.

4.12 $k = 1$ instanton solution for Euclidean $SU(3)$ Yang-Mills theory

In this section, the precise value of the coupling constant does not play any role so for simplicity we will set it to one. The $k = 1$ instanton solution for the $SU(3)$ gauge group in 4-dimensional Euclidean Yang-Mills has not been studied as well as the $k = 1$ instanton solution for the $SU(2)$ gauge group. In this section we will construct the solution for the $SU(3)$ gauge group by exploring the geometrical structure of the S^3 manifold embedded in $SU(3)$. This geometrical structure is naturally reflected in the algebraic structure of the Lie algebra. Our construction is novel. Once we had completed this analysis, we discovered a closely related discussion in [28, p. 13-14]. Our construction emphasizes the geometry more than [28, p. 13-14] does. The two discussions are in complete agreement.

Before entering into the details of our computation, it is useful to briefly recall the logic and main points of our discussion of Yang-Mills instantons. We started by writing the Euclidean action for 4-dimensional Yang-Mills theory. We then found that, to construct solutions of finite action, we need to find field configurations that are pure gauge fields on S^3_∞. Formulating the problem in this way, we were able to recast it as the mathematical problem of finding maps that are classified by the the third fundamental group $\pi_3(G)$ for the gauge group G which concretely amounts to classifying all the mappings: $S^3_\infty \to G$. In this way homotopy classes make a natural appearance and we found that each such homotopy class is equivalent to a given finite action configuration. We further simplified our problem and learned that we only needed to find selfdual (or anti-selfdual) field configurations.

Our discussion for $SU(3)$ will follow the same ideas. Consequently, we want to find $\pi_3(SU(3))$, which will then enable us to determine the set of selfdual field configurations that have finite action, $S_o = 8\pi^2$. It is these finite action configurations that will dominate the semi-classical limit. Our discussion for the $SU(2)$ gauge group showed that $\pi_3(SU(2)) \sim \mathbb{Z}$. This is easily understood if we remember that the geometrical structure of $SU(2)$ is S^3 itself. We are then studying mappings from spheres to spheres which are classified by the winding number, which takes values in $\mathbb{Z}$.

Geometrically $SU(3)$ is a connected and compact 8-dimensional manifold $\mathcal{M}_8$. Bott has carefully studied $\pi_3(SU(3))$ and has concluded that $\pi_3(SU(3)) \sim \mathbb{Z}$. Intuitively we can understand Bott's result by recalling that $SU(2)$ is a subgroup of $SU(3)$. This implies that S^3 is a submanifold of $\mathcal{M}_8$. The maps which are being classified by $\pi_3(SU(3))$ are maps from S^3 into precisely these S^3 submanifolds. The following figure illustrates the above argument.

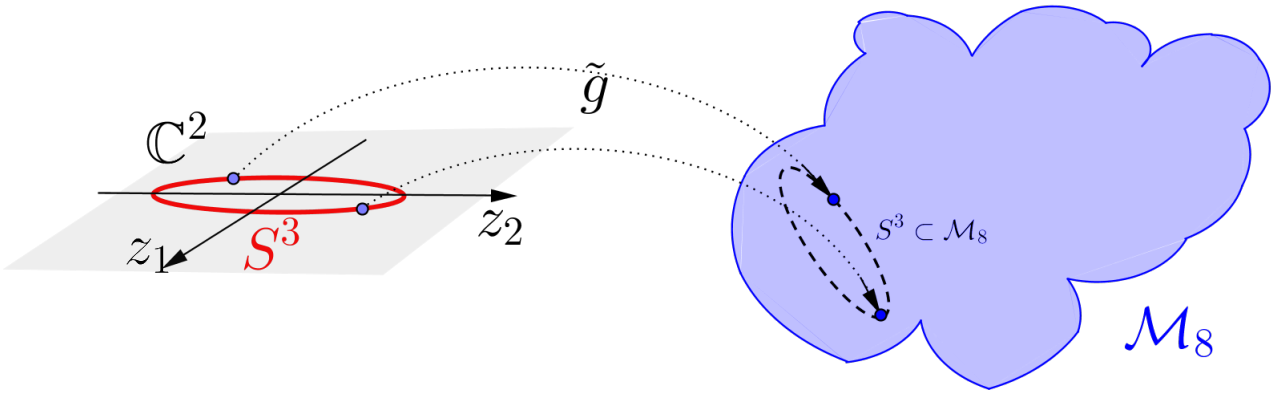


Figure 4.8: $\tilde{g}$ is a map from S^3_∞ to $\mathcal{M}_8$. $\tilde{g}$ wraps a submanifold $S^3 \subset \mathcal{M}_8$ with S^3_∞.

The above discussion is needed to demonstrate completeness of our construction for $k = 1$ instanton solutions for $SU(3)$.

It is useful at this point, to review the $\mathfrak{su}(3)$ Lie algebra. Our $k = 1$ instanton solution will make use of properties of $\mathfrak{su}(3)$. In addition, the above geometrical arguments will become very concrete when understood in terms of the properties of $\mathfrak{su}(3)$. Recall that in general, for $SU(n)$, there are $n^2 - 1$ real degrees of freedom. Accordingly, $SU(3)$ has 8 degrees of freedom. This implies that the dimension of the $\mathfrak{su}(3)$ Lie algebra is 8. A convenient basis for $\mathfrak{su}(3)$ is provided by the Gell-mann matrices. The Gell-mann matrices are

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$
$$\lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

It is straightforward to see that the above matrices satisfy the following defining properties of $\mathfrak{su}(3)$

$$\lambda_i^\dagger = \lambda_i, \quad \text{where} \quad \forall i \in I_8 \equiv \{1, 2, 3, 4, 5, 6, 7, 8\},$$
$$\text{tr}(\lambda_i)|_{\forall i \in I_8} = 0.$$

Often the factor $\frac{1}{\sqrt{3}}$ is omitted. We include this factor to ensure the orthogonality $\text{tr}(\lambda_i \lambda_j)|_{\forall i \in I_8} = 2\delta_{ij}$. Finally, it can been shown by direct calculation that these matrices satisfy the following algebra.

$$[\lambda_a, \lambda_b] = i \sum_k f_{abc} \lambda_c \quad \forall a, b, c \in I_8, \tag{4.12.1}$$

where f_{abc}, called the structure constants, are completely antisymmetric. The only non-zero structure constants are

$$f_{123} = 2, \tag{4.12.2}$$
$$f_{147} = -f_{156} = f_{246} = f_{257} = f_{345} = -f_{367} = 1, \tag{4.12.3}$$
$$f_{458} = f_{678} = \sqrt{3}. \tag{4.12.4}$$

We call $\mathcal{B} = \{\lambda_i\}_{\forall i \in I_8}$ the canonical basis of $\mathfrak{su}(3)$. $\mathcal{B}$ is an orthogonal basis. It is straightforward to see that $\mathcal{B}' = \{\lambda_1, \lambda_2, \lambda_+, \lambda_4, \lambda_5, \lambda_6, \lambda_7, \lambda_-\}$, where $\lambda_\pm = \pm\lambda_3 + \sqrt{3}\lambda_8$, also forms an orthogonal basis of $\mathfrak{su}(3)$. Both of these bases $\mathcal{B}$ and $\mathcal{B}'$ are relevant for the construction of $k = 1$ instantons for $SU(3)$. The sets $\{\lambda_1, \lambda_2, \lambda_3\}$, $\{\lambda_4, \lambda_5, \lambda_+\}$ and $\{\lambda_6, \lambda_7, \lambda_-\}$ form three independent $SU(2)$ subgroups of $SU(3)$. The abstract geometrical construction of an S^3 submanifold of $\mathcal{M}_8$ becomes clear in terms of the set of triplets λ_i's. Indeed, the tangent directions of three different submanifolds $S^3 \subset \mathcal{M}_8$ are those triplets that we have just defined. Each triplet of λ_i's forms a set of three different instanton flavors.

To conclude our preparation for the construction of the $k = 1$ instanton, we need to review 't Hooft selfdual and antisymmetric ansatz $\sigma_{\mu\nu}$ for $SU(2)$. Towards this end, consider the following corollary.

Corollary: The only antisymmetric and selfdual 4×4 matrix solution to $M_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}M_{\rho\sigma}$ is

$$M_{\mu\nu} = \begin{pmatrix} 0 & a & -b & c \\ -a & 0 & c & b \\ b & -c & 0 & a \\ -c & -b & -a & 0 \end{pmatrix}_{\mu\nu}, \tag{4.12.5}$$

where there is no restriction on the set that a, b and c belong to. Since there are three free parameters this is the general form for a selfdual 4×4 matrix.

Now, with the help of this corollary and the previous discussion plus specifically the properties of the constant structure f_{123}, it is trivial to see that

$$\lambda_{\mu\nu} = \frac{1}{2}\begin{pmatrix} 0 & \lambda_3 & -\lambda_2 & \lambda_1 \\ -\lambda_3 & 0 & \lambda_1 & \lambda_2 \\ \lambda_2 & -\lambda_1 & 0 & \lambda_3 \\ -\lambda_1 & -\lambda_2 & -\lambda_3 & 0 \end{pmatrix}_{\mu\nu} \tag{4.12.6}$$

is a selfdual and antisymmetric 4×4 matrix which lives in $\mathfrak{su}(3)$. The decomposition of $\lambda_{\mu\nu}$ in the basis $\mathcal{B}$ is given by $\lambda_{\mu\nu} = \frac{1}{2}(\eta_1, \eta_2, \eta_3, 0, 0, 0, 0, 0)$, where η_i are defined in (4.11.9). Accordingly, by analogy with (4.11.7), the $k = 1$ instanton solution in $SU(2)$, we obtain the $k = 1$ instanton solution for $SU(3)$

$$\mathcal{A}_\mu = \frac{\rho^2(x-X)_\nu}{(x-X)^2[(x-X)^2+\rho^2]} g\lambda_{\mu\nu}g^{-1}. \tag{4.12.7}$$

ρ, X are respectively the size and center of instanton, $g \in SU(3)$ is a global gauge transformation in the directions λ_1, λ_2 and λ_3. Now define $\lambda'_\pm = \frac{1}{2}\lambda_\pm$. It follows that the $k = 1$ instanton solutions for the other two independent triplets, associated to $\{\lambda_4, \lambda_5, \lambda'_+\}$ and $\{\lambda_6, \lambda_7, \lambda'_-\}$, are respectively

$$\mathcal{A}^+_\mu = \frac{\rho^2(x-X)_\nu}{(x-X)^2[(x-X)^2+\rho^2]} g_+\lambda^+_{\mu\nu}g_+^{-1}, \tag{4.12.8}$$

$$\mathcal{A}^-_\mu = \frac{\rho^2(x-X)_\nu}{(x-X)^2[(x-X)^2+\rho^2]} g_-\lambda^-_{\mu\nu}g_-^{-1}. \tag{4.12.9}$$

The above $\lambda^\pm_{\mu\nu}$ are defined as

$$\lambda^+_{\mu\nu} = \frac{1}{2}\begin{pmatrix} 0 & \lambda'_+ & -\lambda_5 & \lambda_4 \\ -\lambda'_+ & 0 & \lambda_4 & \lambda_5 \\ \lambda_5 & -\lambda_4 & 0 & \lambda'_+ \\ -\lambda_4 & -\lambda_5 & -\lambda'_+ & 0 \end{pmatrix}_{\mu\nu}, \quad \lambda^-_{\mu\nu} = \frac{1}{2}\begin{pmatrix} 0 & \lambda'_- & -\lambda_7 & \lambda_6 \\ -\lambda'_+ & 0 & \lambda_6 & \lambda_7 \\ \lambda_7 & -\lambda_6 & 0 & \lambda'_- \\ -\lambda_6 & -\lambda_7 & -\lambda'_- & 0 \end{pmatrix}_{\mu\nu}. \tag{4.12.10}$$

The global gauge transformation $g_{\pm}$ are respectively performed in the directions $\{\lambda_4, \lambda_5, \lambda'_+\}$ and $\{\lambda_6, \lambda_7, \lambda'_-\}$. These results are simply explained by noting that the three triplets $\{\lambda_1, \lambda_2, \lambda_3\}$, $\{\lambda_4, \lambda_5, \lambda'_+\}$ and $\{\lambda_6, \lambda_7, \lambda'_-\}$ share the same algebra $[\lambda_1, \lambda_2] = if_{123}\lambda_3$. It is straightforward to see that the components of $\lambda^{\pm}_{\mu\nu}$ in the basis $\mathcal{B}$ are respectively

$$\lambda^{+}_{\mu\nu} = \frac{1}{2}(0, 0, \frac{\eta_3}{2}, \eta_1, \eta_2, 0, 0, \frac{\sqrt{3}}{2}\eta_3),$$
$$\lambda^{-}_{\mu\nu} = \frac{1}{2}(0, 0, -\frac{\eta_3}{2}, 0, 0, \eta_1, \eta_2, \frac{\sqrt{3}}{2}\eta_3).$$

The solutions given in (4.12.7), (4.12.8) and (4.12.9) are the solutions for a $k=1$ instanton for gauge group $SU(3)$ in 4-dimensional Euclidean space. These solutions are inequivalent in the sense that the zero modes coming from the global gauge transformations are different. Having these solutions, we can extend our investigation of the moduli space of the $k=1$ instanton solution. Indeed, with the help of the collective coordinates $\chi_\mu = (X_1, X_2, X_3, X_4, \rho, \theta_1, \ldots, \theta_8)$ one can show that the dimension of the moduli space for $k=1$ instanton solution for $SU(3)$ is $D=12$. This is due to the fact that the relation $[\lambda_4, \lambda_5] - i\lambda_3 = [\lambda_6, \lambda_7] + i\lambda_3$ relates the parameters $\theta_3, \ldots, \theta_8$. The 12 zero modes are in perfect agreement with the zero modes counting in [3]. The counting in [3] is based on the index theorem and concludes that the number of zero modes for k instantons in $SU(n)$ is $4kn$. In our case we have $k=1, n=3$ thus $D=12$ is indeed the number of zero modes expected.

To conclude this section, note that we already have a clue for the construction of the $k=1$ instanton solution for $SU(n)$. In fact, among the n^2-1 basis vectors of $\mathfrak{su}(n)$, the first three vectors of the Lie algebra always have the form

$$T_1 = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 0 \end{pmatrix}, \quad T_2 = \begin{pmatrix} 0 & -i & 0 & \ldots & 0 \\ i & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 0 \end{pmatrix}, \quad T_3 = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & -1 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 0 \end{pmatrix}.$$

These matrices satisfy the commutation relation $[T_1, T_2] = if_{123}T_3$, where again $f_{123}=2$ is completely antisymmetric. Just like in the previous discussions, these matrices form a submanifold S^3 of the geometrical structure of $SU(n)$. Thus, we find one type of $k=1$ instanton solution

$$\mathcal{A}_\mu = \frac{\rho^2(x-X)_\nu}{(x-X)^2[(x-X)^2+\rho^2]} g\mathcal{T}_{\mu\nu}g^{-1}, \tag{4.12.11}$$

where $g \in SU(n)$ is a global gauge transformation along the directions T_1, T_2 and T_3 while $\mathcal{T}_{\mu\nu}$ is defined by

$$\mathcal{T}_{\mu\nu} = \begin{pmatrix} 0 & T_3 & -T_2 & T_1 \\ -T_3 & 0 & T_1 & T_2 \\ T_2 & -T_1 & 0 & T_3 \\ -T_1 & -T_2 & -T_3 & 0 \end{pmatrix}. \tag{4.12.12}$$

By noting that $SU(3)$ is also a subgroup of $SU(n)$ when $n>3$, we can start by inserting in the left upper corner of T_i the others λ_i to yield further $k=1$ solutions for $SU(n)$ in general.

5. $D = 5$ $SU(2)$ Maximally Supersymmetric Yang-Mills Theory (MSYM) and the $\mathcal{N} = (2, 0)$ theory

In this chapter we will return to a general gauge coupling λ and we will return to Minkowski space. We will explain the relevance of instanton solutions to the $D = 4$ dimensional Euclidean Yang-Mills theory for the $D = 5$ Minkowski space Yang-Mills theory. Moreover, we will derive dyonic instantons after adding to the pure bosonic Yang-Mills theory a complex scalar potential. $D = 5$ dimensional super Yang-Mills theory arises as the low energy dynamics of open strings attached to $D4-$branes. The $D4-$branes set up Dirichlet boundary conditions for the endpoints of the open strings in the directions orthogonal to the $D4-$branes and Neumann boundary conditions in the directions parallel to the $D4-$branes. The position of the open string endpoints in directions parallel to the $D4-$branes is described by the Yang-Mills gauge field, while the position of the open string endpoints in the directions orthogonal to the $D4-$branes are described by scalar fields transforming in the adjoint representation of the gauge group. In this way, the spectrum of open string excitations for open strings stretched between different $D4-$branes contain the $U(N)$ gauge fields $\mathcal{A}_\mu$ where μ runs from 0 up to 4, and five scalar fields. In our study we consider a stack of two $D4-$branes so that the gauge group is $SU(2)$.

As described earlier, due to the boundary of the worldsheet, the left- and right-moving fields are related. This breaks half of the allowed supersymmetries so that the theory in the presence of a stack of $D4-$branes preserves half of the allowed supersymmetries of the full spacetime symmetry group. Therefore this theory will enjoy an $\mathcal{N} = (2, 0)$ supersymmetry in 5 dimensions. The instantons are $1/2-$BPS meaning that these solutions only enjoy half of the (2,0) supersymmetry. Finally, when the scalar fields are excited to obtain dyonic instantons, the solutions preserve only one half of the supersymmetry preserved by the instanton. We summarize this by saying that the dyonic instanton is a $1/4-$BPS object.

5.1 Instantons

In the previous chapter, we have seen that instantons are solutions to the pure bosonic Euclidean Yang-Mills theory in four dimensions. Thus it is natural to consider only the bosonic fields of the $D = 5$ Yang-Mills theory. The bosonic action is

$$S = -\frac{1}{4\lambda^2}\int d^5x \mathrm{tr}(\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu}). \tag{5.1.1}$$

We can see from this action that in 5 dimensions the dimension of the gauge coupling is $[\lambda] = L^{-\frac{1}{2}}$. This means that Yang-Mills interaction is irrelevant, or equivalently, that the theory is non-renormalizable. Note that we are now in Minkowski space so that the contraction of the indices is with the usual Minkowski metric. In addition, the covariant derivative, the gauge fields and the field strengths must be Hermitian. Concretely, our conventions are

$$\begin{aligned} D_\mu &= \partial_\mu - i\lambda\mathcal{A}_\mu, \\ \mathcal{F}_{\mu\nu} &= \partial_\mu\mathcal{A}_\nu - \partial_\nu\mathcal{A}_\mu - i\lambda[\mathcal{A}_\mu, \mathcal{A}_\nu]. \end{aligned}$$

Having the above action, by analogy to the equation (3.3.20), it is straightforward to see that the energy of the system is

$$E = \frac{1}{2\lambda^2}\int d^4x \mathrm{tr}(\mathcal{F}_{i0}\mathcal{F}_{i0} + \mathcal{F}_{ij}\mathcal{F}_{ij}). \tag{5.1.2}$$

Since $\mathrm{tr}(\mathcal{F}_{i0}\mathcal{F}_{i0}) > 0$, a calculation using the ideas that give (4.5.14) can be performed to show that

$$E \geq \frac{4\pi^2|k|}{\lambda^2}, \tag{5.1.3}$$

where k is the instanton number

$$k = \frac{-1}{32\pi^2}\int d^4x \epsilon_{ijkl}\mathcal{F}_{ij}\mathcal{F}_{kl}. \tag{5.1.4}$$

The above inequality proves that instantons are solutions to the low energy dynamics of the $D = 5$ MSYM. Since the instanton numbers are integers these minima correspond to discrete energy values.

5.2 Dyonic instantons

As described above, dyonic instantons arise after adding complex scalar fields $\phi^a(x)$ to the Lagrangian (5.1.1). The presence of a non-zero complex scalar field does not affect the instanton field configurations. As a result dyonic instantons inherit the geometrical properties of instantons. Indeed, dyonic instantons are also solutions to the selfdual field equation but now with a non-zero scalar potential. The scalar potential ϕ^a encodes the separations of the $D4-$branes in the directions transverse to the $D4-$branes. ϕ^a lives in $\mathfrak{su}(2)$ so that $[A_\mu, \phi] \neq 0$. For simplicity, we will assume that the only non-zero VEV is in the x_4 direction so that the dyonic instanton action is

$$S = \int d^5x \mathrm{tr}\left(-\frac{1}{4\lambda^2}\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu} + \frac{1}{2}D_\mu\phi^\star D^\mu\phi\right) \tag{5.2.1}$$

Using this action for the dyonic instantons, the corresponding energy is

$$\begin{aligned} E &= \frac{1}{2}\int d^4x \mathrm{tr}\left(\tfrac{1}{\lambda^2}(\mathcal{F}_{i0})^2 + \tfrac{1}{\lambda^2}(\mathcal{F}_{ij})^2 + (D_0\phi)^2 + (D_i\phi)^2\right), \\ &= \frac{1}{2}\int d^4x \mathrm{tr}\left((\tfrac{1}{\lambda}\mathcal{F}_{i0} - D_i\phi)^2 + \tfrac{1}{4\lambda^2}(\mathcal{F}_{ij} - \tilde{\mathcal{F}}_{ij})^2 + (D_0\phi)^2 + \tfrac{1}{2\lambda^2}\mathcal{F}_{ij}\tilde{F}_{ij} + 2\tfrac{1}{\lambda}\mathcal{F}_{i0}D_i\phi\right) \end{aligned}$$

In term of the dual of the field strength, $\tilde{\mathcal{F}}_{ij} = \frac{1}{2}\epsilon_{ijkl}\mathcal{F}_{kl}$, we can rewrite the energy as

$$E = \frac{1}{2}\int d^4x \mathrm{tr}\left((\tfrac{1}{\lambda}\mathcal{F}_{i0} - D_i\phi)^2 + \tfrac{1}{4\lambda^2}(\mathcal{F}_{ij} - \tilde{\mathcal{F}}_{ij})^2 + (D_0\phi)^2 + \tfrac{1}{2\lambda^2}\mathcal{F}_{ij}\tilde{F}_{ij} + 2\tfrac{1}{\lambda}\mathcal{F}_{i0}D_i\phi\right). \tag{5.2.2}$$

From this new expression of the energy, we see that the conditions

$$\mathcal{F}_{ij} = \tilde{\mathcal{F}}_{ij}, \tag{5.2.3}$$
$$\mathcal{F}_{i0} = D_i\phi, \tag{5.2.4}$$
$$D_0 = 0; \tag{5.2.5}$$

attain the lower bound of the energy, given by

$$E \geq \frac{8\pi^2|k|}{\lambda^2} + |Q_E|, \tag{5.2.6}$$

where k is the usual instanton number and Q_E is the electrical charge defined by

$$Q_E = \int d^4x \text{tr}(D_i\phi)^2. \tag{5.2.7}$$

The inequality (5.2.6) tells us that dyonic instantons are solutions of minimum energy. (5.2.3) is the condition for a selfdual field strength. Together the equations (5.2.3), (5.2.4) and (5.2.5) are the BPS equations for the dyonic instanton. The second and third equations are satisfied when the spatial gauge fields are static and $\mathcal{A}_0 = \phi$. Indeed, we have

$$\begin{cases} \mathcal{F}_{i0} = D_i\phi \\ D_0\phi = 0 \end{cases} \Leftrightarrow \begin{cases} \partial_i\mathcal{A}_0 - \partial_0\mathcal{A}_i - i\lambda[\mathcal{A}_i, \mathcal{A}_0] = \partial_i\phi - i\lambda[\mathcal{A}_i, \phi] \\ \partial_0\phi - i\lambda[A_0, \phi] = 0 \end{cases} \Rightarrow \begin{cases} \partial_0\mathcal{A}_i = 0 \\ A_0 = \phi. \end{cases}$$

The equation of motion for the scalar field

$$D_\mu D^\mu\phi = 0 \tag{5.2.8}$$

still needs to be solved. In cases where there is more than one VEV we need to solve for each of the corresponding scalar fields. When the VEV of ϕ vanishes, ϕ itself vanishes everywhere and the dyonic instantons reduce to the neutral particle-like instantons.

We know that the $k > 1$ instantons make an exponentially small contribution to the path integral (see (4.5.20)), and hence their contributions are negligible. Similarly, it is enough, with exponential accuracy, to consider only the single dyonic instanton. The $k = 1$ dyonic instanton solutions for $SU(2)$ are [29]

$$\mathcal{A}_i = \frac{\frac{2}{\lambda}\rho^2 x_j}{x^2[x^2 + \rho^2]}\eta_{aij}\frac{\sigma_a}{2}, \quad \phi = \frac{x^2}{x^2 + \rho^2}\frac{\sigma_3}{2}. \tag{5.2.9}$$

The spatial components of the gauge field are the $k = 1$ instanton solution for $SU(2)$ as we claimed. This single instanton solution is centered at $X = 0$ and the gauge zero modes are not exhibited in the solution. Using the above solution, one computes the field strength $\mathcal{F}_{i0} = D_i\phi$ to find

$$D_i\phi = \mathcal{F}_{i0} = \frac{2\rho^2 x_i}{(x^2 + \rho^2)^2}\eta_{3jl}\eta_{ail}\frac{\sigma_a}{2}. \tag{5.2.10}$$

This expression for the temporal field strength allows us to evaluate the electric charge of the dyonic instanton. Indeed, using the formula (5.2.7), one finds [29]

$$Q_E = \int d^4x \text{tr}(\mathcal{F}_{i0}\mathcal{F}_{i0}) = 4\pi^2\rho^2. \tag{5.2.11}$$

Finally, this equation demonstrates that an instanton of size $\rho \approx \sqrt{|Q_E|}$ is stabilized by the electric charge of the dyonic instanton. In other words the fact that we coupled the scalar potential field to the pure instantonic field has produced a stable and static instanton.

5.3 Instantons in $D = 5$ MSYM and the $\mathcal{N} = (2, 0)$ theory

The $\mathcal{N} = (2, 0)$ theory is a 6-dimensional superconformal quantum field theory. It has been conjectured that the compactification on a circle, (see Figure 5.1) of the $\mathcal{N} = (2, 0)$ theory gives a theory, that at low energy is equivalent to $D = 5$ MSYM. The $D = 5$ MSYM contains 16 supercharges.

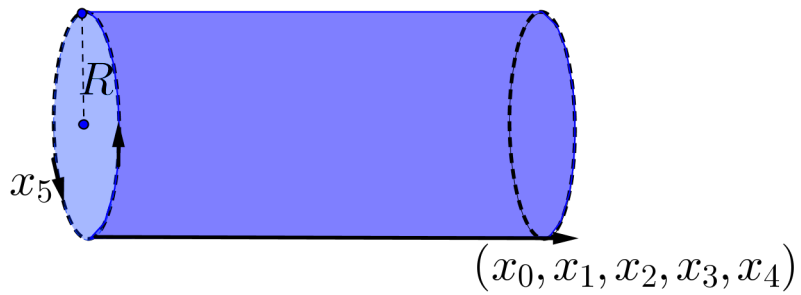


Figure 5.1: Figure illustrating the compactification on a circle of a $6-$dimensional Minkowski spacetime.

The above figure illustrates that the compactification is constructed by performing an identification in the x_5 direction. Given the coordinates $(x_o, x_1, x_2, x_3, x_4, x_5)|_P$ for a point P, for any integer k the point with coordinates $(x_o, x_1, x_2, x_3, x_4, x_5 + 2k\pi R)$ is identified as the same point P. In other words, we simply write $x_5 \sim x_5 + 2k\pi R$, where $k \in \mathbb{Z}$ and R is the radius of the circle in the figure above.

We demonstrated in the previous section that solutions to the low energy dynamics of the $D = 5$ MSYM can be obtained from the the instanton solution to the four dimensional Euclidean Yang-Mills theory. Accordingly, we will show that if $D = 5$ MSYM is equivalent to the compactification on a circle of the $\mathcal{N} = (2, 0)$ theory, instantons in the lower dimensional theory must be identified with the Kaluza-Klein modes of the compactified dimension. This statement suggests that instantons are lifted to real particles embedded in the extra dimensions.

Now, return to the relation (5.1.3). It is straightforward to see that at the lowest energy scale of the theory, the gauge coupling of the $D = 5$ MSYM is

$$\lambda^2 = \frac{4\pi^2}{E} \equiv \frac{4\pi^2}{M_{\text{inst}}}. \tag{5.3.1}$$

The Einstein equivalence between mass-energy has been used to identify that the quantity M_{inst} as the mass of a single instanton. Following the discussions in [30, 31, 32], it is known that the size of the compactified dimension is given by $R \sim \lambda^2$ where R is the radius of the circle of compactification. Also, it is known [3] that the Kaluza-Klein modes (KK-modes), corresponding to momentum modes around the circle, have mass $M_{KK} = \frac{1}{R}$. Hence, in summary we find

$$M_{KK} = M_{\text{inst}} = \frac{1}{R}. \tag{5.3.2}$$

Accordingly at very strong coupling $\lambda^2 \to \infty$ the mass of the instanton goes to zero. Since an instanton of arbitrary size is a stable particle in the presence of a scalar field the behaviour of instanton mass is the basis for the claim that instantons are real particles in the extra dimension.

6. Conclusions

This work has achieved a rather broad review of the physical applications of instanton solutions. Indeed, applications ranging from the simple example of a potential in one dimensional quantum mechanics to complicated quantum field theories, including supersymmetric Yang-Mills theory have been described. In all of these examples, the vacuum to vacuum transition amplitude has been computed using a semi-classical expansion. These expansions are always dominated by the classical Euclidean instanton solutions. In other words, instantons play an important role in determining the vacuum structure of the theory.

This physical relevance of instantons may at first sight be surprising, given that the instantons are solutions to the equations of motion derived from the Euclidean action, obtained after performing a Wick rotation of the time direction. One might have expected that performing a Wick rotation of the time will divorce us from the physics of the original theory. However, in the case of quantum mechanics we were able to exactly recover the even part of the energy spectrum of the simple harmonic oscillator, directly from the Euclidean space solutions. Moreover, our comparison of the WKB solution to the Schrödinger equation for the double well potential and the instanton solution to the double well potential are in complete accord. Further, we showed that instanton solutions for the periodic potential correctly lead the continuum band of energy states predicted by electronic band theory of solid state physics. Together, this is a convincing demonstration of the physical relevance of the Euclidean theory.

One technical detail that we used, is that the contribution to the transition amplitude of the small fluctuations around a configuration of n well separated instantons is given by K^n, where K is the contribution of the small fluctuations around a single instanton. Although this is a standard assumption, there is no clear technical justification for it in the literature as far as we can see. We have demonstrated this fact by showing that the number K is given by the zero mode measure times the inverse square root of the product of the energies of any boundstates of a well in V'' and the effect of the well on the continuum eigenvalues of a partial differential equation defining the instanton zero modes. Each additional instanton in the multi-instanton configuration supplies an extra potential well and hence an extra factor of the bound state eigenvalues.

Our study of instantons in quantum mechanics was primarily to prepare for a study of instantons in quantum field theory. In the quantum field theory, a study of instantons leads not only to interesting physics but also to fascinating mathematics. Indeed, we have learned that instantons are field configurations that are pure gauge on the surface of the three sphere at infinity, in 4-dimensional Euclidean space. Based on this, we argued that instantons are classified by the third fundamental group $\pi_3(G)$ of the gauge group G. Our results imply that we can realize $\pi_3(G)$ in terms of the selfdual solutions to the field equations. With 't Hoofts ansatz for the $k = 1$ instanton solution for $SU(2)$ in hand and with the help of the ADHM construction, which demonstrates the existence of the $k > 1$ instanton solutions for any gauge group, we were able to construct completely the $k = 1$ instanton solution for $SU(3)$. Further, we have introduced a particular for $k = 1$ instanton solution for $SU(n)$, for any n. The $k = 1$ instanton solution for $SU(n)$ is easily constructed following our algorithm for $SU(3)$.

Our conventional understanding is that instantons do not represent real particles, like for example, the photons of QED. Indeed, they are solutions to the Euclidean theory obtained by analytic continuation of the Minkowski theory. Instantons do however encode non-trivial information about the low lying energy states of the theory. In contrast to this conventional wisdom, as we have reviewed, instantons are solutions to the low energy dynamics of the $D = 5$ Minkowski MSYM theory. These instanton solutions are not stable unless we add extra scalar potentials to the pure Yang-Mills action. The extra

scalar potentials arise very naturally in the low energy theory of the open strings attached to $D4-$branes. In this setting, the $D = 5$ Minkowski MSYM theory might provide a description of the dynamics of a compactified six dimensional theory describing the physics of an M-theory fivebrane. In this setting, the instantons behave like real particles of the six dimensional theory that have momentum along the compactified dimension. Of course this proposal is very speculative. However, our review of this idea, based on [30, 31, 32, 3], convinces us that this idea deserves further study.

Appendix A. Numerical Spectrum for the double well potential

A.1 Numerical computation of K for the double well potential

Here our goal is to find the eigenvalues determined by the eigenproblem (2.7.27) and then compute K using it expression given by (2.7.26). We start by choosing a suitable value for T, which defines the domain of t. To this end, consider the analytic solution $\tilde{x}(t)$ for a single instanton,

$$\tilde{x}(t) = \frac{\tanh(t - t_2)}{\sqrt{2g}}. \tag{A.1.1}$$

T must be chosen large enough that

$$\frac{1}{\sqrt{2g}} = \frac{\tanh(\frac{T}{2} - t_2)}{\sqrt{2g}} \;\Leftrightarrow\; 1 \approx \tanh(\frac{T}{2} - t_2). \tag{A.1.2}$$

If we choose $\frac{T}{2} = 10$ and $t_2 = 0$, the error between $\tanh(10)$ and 1 is already less than 4×10^{-9}. Thus, setting $T = 20$ is a good approximation for our numerical computation. The eigenvalues are obtained by searching for the values of λ at which the functions $\psi_\lambda(t)$ obey the boundary conditions

$$\psi_{\lambda_i}(-10) = 0, \quad \left.\frac{\partial \psi_{\lambda_i}(t)}{\partial t}\right|_{t=-10} = 1; \quad \text{and} \quad \psi_{\lambda_i}(10) = 0. \tag{A.1.3}$$

We use the first two boundary conditions above together with the ode113() function in matlab to numerically construct the function $\psi_{\lambda_i}(t)$. We then test the third condition above. Whenever it is obeyed, we know that we have found an eigenvalue. The explicit form of the equation that we want to solve is

$$\left(-\frac{d^2}{dt^2} + 6\tanh^2(t - t_2) - 2\right)\psi_{\lambda_i} = \lambda_i \psi_{\lambda_i}. \tag{A.1.4}$$

In order to implement the problem numerically, we need to convert it to first order form. This is easily achieved as follows

$$\frac{d\psi_{\lambda_i}}{dt} = \chi(t), \tag{A.1.5}$$

$$\frac{d\chi(t)}{dt} = (6\tanh^2(t - t_2) - 2 - \lambda_i)\psi_{\lambda_i}, \tag{A.1.6}$$

which can also be written as

$$\frac{d}{dt}\begin{pmatrix} \psi_{\lambda_i}(t) \\ \chi(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 6\tanh^2(t - t_2) - 2 - \lambda_i & 0 \end{pmatrix}\begin{pmatrix} \psi_{\lambda_i}(t) \\ \chi(t) \end{pmatrix}. \tag{A.1.7}$$

In the figure below we have plotted the numerical results for typical choices for the eigenvalue E.

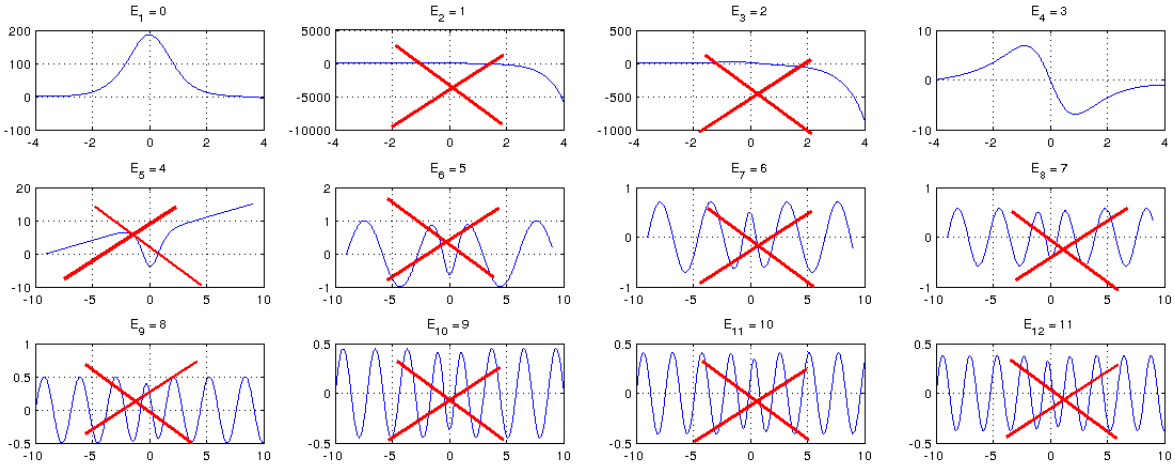


Figure A.1: Eigenfunctions which corresponds to the eigenvalues $\lambda = E_i$.

Figure A.1 shows that $\lambda_o = 0$ and $\lambda_1 = 3$ are eigenvalues of (2.7.27). A careful search over λ shows that these are the only eigenvalues. These results are in perfect agreement with the analytic solutions of (2.7.27). Thus, we find

$$K = \sqrt{\frac{S_o}{2\pi\hbar}}\frac{1}{\sqrt{3}}\sqrt{9} = \sqrt{\frac{1}{\pi g\hbar}}. \tag{A.1.8}$$

References

[1] Michael R. Douglas. On D=5 super Yang-Mills theory and (2,0) theory. *JHEP*, 1102:011, 2011, arXiv/1012.2880 [hep-th].

[2] James P. Allen and Douglas J. Smith. The low energy dynamics of charge two dyonic instantons. 2012, arXiv/1210.3208 [hep-th].

[3] David Tong. TASI lectures on solitons: Instantons, monopoles, vortices and kinks. 2005, arXiv/0509216 [hep-th].

[4] R. Rajaraman. *Solitons and Instantons: An Introduction to Solitons and Instantons in Quantum Field Theory*. North-Holland personal library. North-Holland Publishing Company, 1982.

[5] Michael E. Peskin and Dan V. Schroeder. *An Introduction To Quantum Field Theory (Frontiers in Physics)*. Westview Press, 1995.

[6] J. Polchinski. *String Theory: Volume 1, Superstring Theory and Beyond*. String Theory 2 Volume Hardback Set. Cambridge University Press, 2001.

[7] J. Polchinski. *String Theory: Volume 2, Superstring Theory and Beyond*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2005.

[8] Joachim Ankerhold. *Quantum Tunneling in Complex System, The Semiclassical Aproch*. Springer-Verlag, Berlin, Heidelberg, 2007.

[9] M. Nakahara. *Geometry, Topology, and Physics*. Graduate Student Series in Physics. Institute of Physics Pub., 2003.

[10] R.W. Butler. *Saddlepoint Approximations with Applications*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press., 2007.

[11] P. Deift and X. Zhou. A steepest descent method for oscillatory riemann-hilbert problems. asymptotics for the mkdv equation. *Annals of Mathematics*, 137(2):pp. 295–368, 1993.

[12] A.I. Vainshtein, Valentin I. Zakharov, V.A. Novikov, and Mikhail A. Shifman. ABC's of Instantons. *Sov.Phys.Usp.*, 25:195, 1982.

[13] Bahram Houchmandzadeh. *Mathématiques pour la Physique*. Univerité de Grenoble, Département de Physique, 2009.

[14] L. D. Landau and E. M. Lifshitz. *QUANTUM MECHANICS: Non-relativistic Theory*. Pergamon Press., 1977.

[15] Sidney Coleman. *Aspects of Symmetry: Selected Erice Lectures*. Cambridge University Press, February 1988.

[16] P. A. M. Dirac. Generalized hamiltonian dynamics. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 246(1246):pp. 326–332, 1958.

[17] A.A. Belavin, Alexander M. Polyakov, A.S. Schwartz, and Yu.S. Tyupkin. Pseudoparticle Solutions of the Yang-Mills Equations. *Phys.Lett.*, B59:85–87, 1975.

[18] R. Jackiw and C. Rebbi. Conformal Properties of a Yang-Mills Pseudoparticle. *Phys.Rev.*, D14:517, 1976.

[19] R. Jackiw, C. Nohl, and C. Rebbi. Conformal Properties of Pseudoparticle Configurations. *Phys.Rev.*, D15:1642, 1977.

[20] R. Jackiw and C. Rebbi. Degrees of freedom in pseudoparticle systems. *Physics Letters B*, 67:189–192, 1977.

[21] G. 't Hooft. Computation of the quantum effects due to a four-dimensional pseudoparticle. *Phys. Rev. D*, 14:3432–3450, Dec 1976.

[22] Nick Dorey, Timothy J. Hollowood, Valentin V. Khoze, and Michael P. Mattis. The Calculus of many instantons. *Phys.Rept.*, 371:231–459, 2002, arXiv/0206063 [hep-th].

[23] E. Corrigan and P. Goddard. Construction of instanton and monopole solutions and reciprocity. *Ann. Phys.*, 154:253–279, 1984.

[24] Michael F. Atiyah; Nigel J. Hitchin and I.M. Singer. Self-duality in four-dimensional Riemannian geometry. *Proc. R. Soc. Lond., Ser. A*, 362:425–461, 1978.

[25] A. S. Schwarz. Instantons and fermions in the field of instanton. *Comm. Math. Phys.*, 64(3):233–268, 1979.

[26] Edward Corrigan, P. Goddard, H. Osborn, and S. Templeton. Zeta Function Regularization and Multi - Instanton Determinants. *Nucl.Phys.*, B159:469, 1979.

[27] H. Osborn. Solutions of the Dirac Equation for General Instanton Solutions. *Nucl.Phys.*, B140:45, 1978.

[28] Stefan Vandoren and Peter van Nieuwenhuizen. Lectures on instantons. 2008, 0802.1862.

[29] Neil D. Lambert and David Tong. Dyonic instantons in five-dimensional gauge theories. *Phys.Lett.*, B462:89–94, 1999, arXiv/9907014 [hep-th].

[30] Moshe Rozali. Matrix theory and U duality in seven-dimensions. *Phys.Lett.*, B400:260–264, 1997, arXiv/9702136 [hep-th].

[31] Micha Berkooz, Moshe Rozali, and Nathan Seiberg. Matrix description of M theory on T**4 and T**5. *Phys.Lett.*, B408:105–110, 1997, arXiv/9704089 [hep-th].

[32] Nathan Seiberg. Notes on theories with 16 supercharges. *Nucl.Phys.Proc.Suppl.*, 67:158–171, 1998, arXiv/9705117 [hep-th].

[33] Kenneth Wolsson. A condition equivalent to linear dependence for functions with vanishing wronskian. *Linear Algebra and its Applications*, 116(0):1 – 8, 1989.

[34] Edward Corrigan, D.B. Fairlie, S. Templeton, and P. Goddard. A Green's Function for the General Selfdual Gauge Field. *Nucl.Phys.*, B140:31, 1978.

[35] Norman H. Christ, Erick J. Weinberg, and Nancy K. Stanton. General self-dual yang-mills solutions. *Phys. Rev. D*, 18:2013–2025, Sep 1978.